THE
PHYSICAL REVIEW
A journal of experimental and theoretical physics established by E. L. Nichols in 1893

SECOND SERIES, VOL. 66, NOS. 7 AND 8 OCTOBER 1 AND 15, 1944

Theory of Diffraction by Small Holes

H. A. BETHE
Department of Physics, Cornell University, Ithaca, New York
(Received January 26, 1942)

The diffraction of electromagnetic radiation by a hole small compared with the wave-length is treated theoretically. A complete solution is found satisfying Maxwell's equations and the boundary conditions everywhere (Section 4). The solution holds for a circular hole in a perfectly conducting plane screen, but it is believed that the method will be applicable to much more general problems (Section 8). The method is based on the use of fictitious magnetic charges and currents in the diffracting hole which has the advantage of automatically satisfying the boundary conditions on the conducting screen. The charges and currents are adjusted so as to give the correct tangential magnetic, and normal electric, field in the hole. The result (Section 5) is completely different from that of Kirchhoff's method, giving for the diffracted electric and magnetic field values which are smaller in the ratio (radius of the hole/wave-length) (Section 6). The diffracted field can be considered as caused by a magnetic moment in the plane of the hole, and an electric moment perpendicular to it (Section 6). The theory is applied to the problem of mutual excitation of cavities coupled by small holes (Section 9). This leads to equations very similar to those for ordinary coupled circuits. The phase and amplitude relations of two coupled cavities are not uniquely determined, but there are two modes of oscillation, of slightly different frequency, for which these relations are opposite (Section 10). The problem of stepping up the excitation from one cavity to another is treated (Section 11).

1. THE PROBLEM

IN microwave work it is often important to know the effect of a small hole in a cavity upon the oscillation of that cavity. For instance, two cavities may be coupled by a small hole in their common boundary (Fig. 1); in this case, we wish to know the characteristic frequencies and the phase relations for the oscillations of the coupled system. Or a hole in a cavity may serve the purpose of getting radiation out of it; then we want to calculate the amount and the spatial distribution of the emitted radiation. Another similar problem would be to calculate the effect of a small gap in a wave guide upon the propagation of waves along that guide.

A less practical problem but probably the simplest one of the same type, is the *diffraction of electromagnetic waves by a small hole in an infinite plane conducting screen.* This is the problem which we are going to solve first

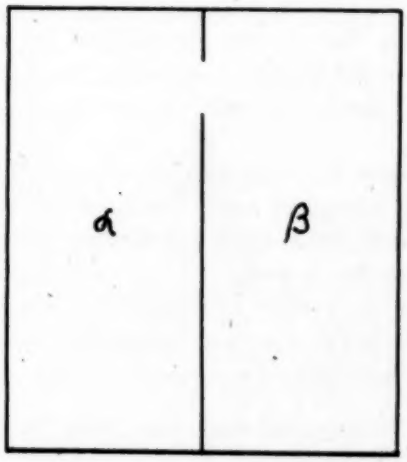

FIG. 1. Two cavities, α and β, coupled by a small hole.

(Sections 3–7); the application of the solution to the practical problems mentioned is then rather straightforward (Sections 9–11).

The available theoretical methods are entirely inadequate for the treatment of our problem. In the usual Kirchhoff method, the diffracted field is expressed in terms of the incident field in the hole. However, the Kirchhoff solution does not satisfy the boundary conditions, *viz.*, it does not give zero tangential component of the electric field on the screen. In most textbooks, the pious hope is expressed that Kirchhoff's method will give at least the first term of a convergent series. This is probably true for the diffraction by an opening, large compared with the wave-length, because then the diffracted field will be relatively small on the screen, thus "almost" fulfilling the boundary conditions. But it is certainly not true for a small hole; in fact, our exact solution of the problem will turn out to be entirely different from Kirchhoff's. The failure of Kirchhoff's theory will be demonstrated mathematically in Section 2.

Kirchhoff's method has the additional defect of being a scalar theory while the electromagnetic field is essentially vectorial. This shortcoming has been remedied in the last forty years by a number of writers; a very good account of the vector equivalent of Kirchhoff's theory is given in Stratton's book,[1] to which we shall frequently refer in this paper. The vectorial theory ensures the fulfillment of the divergence conditions, $\operatorname{div} E = \operatorname{div} H = 0$; i.e., it gives transverse waves in the wave zone which would not necessarily be the case in Kirchhoff's scalar formulation. However, the vector formulation in no way improves the situation regarding the fulfillment of the boundary conditions on the conducting screen.

The only rigorous solution of any diffraction problem known to me is Sommerfeld's solution[2] of the diffraction by a conducting semi-infinite plane, or by a wedge. As is well known, this solution is rather complicated although the problem is the simplest imaginable, being two- rather than three-dimensional. It appears hopeless to look for a rigorous solution of our problem along the lines of Sommerfeld's solution. A slight similarity between his and our method will be pointed out later (Section 6d).

The main simplifying assumption we are going to use is that the hole is small compared with the wave-length. This means that the incident electromagnetic field is almost constant over the hole. We believe, however, that our method can be generalized to holes of larger size (Section 8).

2. THE FAILURE OF KIRCHHOFF'S THEORY

The failure is most easily seen in the simplest formulation of the theory, *viz.*, Kirchhoff's own, scalar formulation. Let u be a scalar wave function satisfying the wave equation

$$\nabla^2 u + k^2 u = 0. \tag{1}$$

We may identify u, e.g., with one of the components of the electric field, say E_y. The derivatives of u will then be connected with the magnetic field. If the conducting screen is placed at $x=0$, the boundary condition is

$$u = 0 \quad \text{at} \quad x = 0. \tag{2}$$

Let electromagnetic waves come in from the left ($x<0$), the corresponding wave function may be u_0. Then according to Green's theorem, the wave function at any point r on the right of the screen is

$$u(r) = \int d\sigma \left[-\frac{\partial u_0}{\partial x'}(r')\varphi(|r-r'|) + u_0(r')\frac{\partial \varphi}{\partial x'} \right] \tag{3}$$

with the Green's function

$$\varphi(r) = e^{ikr}/r. \tag{3a}$$

The integration variables are y', z'. The coordinates r of the "field point" and x' ($=0$) are kept constant. The solution (3) is rigorous if the integral is extended over the entire surface $x'=0$, with the correct values for $u_0(0, y', z')$. However, the distribution of u_0 on the plane $x'=0$ is not known. Kirchhoff's method consists now in putting $u_0 = \partial u_0/\partial x' = 0$ everywhere on the screen outside the hole, and replacing u_0 by the field of the incident wave *in* the hole.

Suppose now the hole is very small compared with the wave-length, then u_0 may be assumed to

[1] J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill Book Company, Inc., New York, 1941), p. 464 ff.

[2] A. Sommerfeld, Riemann-Weber's *Differentialgleichungen der Physik*, seventh edition, p. 433.

be constant over the hole and we obtain, remembering that φ is a symmetrical function of r and r',

$$u(r) = -A\left[\frac{\partial u_0}{\partial x'}\varphi(r) + u_0\frac{\partial \varphi(r)}{\partial x}\right], \qquad (4)$$

where A is the area of the hole. We can now make two alternative assumptions, *viz.*:

(*a*) u_0 in the hole has the value given by the unperturbed incident wave. Then both u_0 and $\partial u_0/\partial x'$ are different from zero. If we have an incident plane wave traveling at an angle ϑ with the x direction, we have

$$\partial u_0/\partial x' = iku_0 \cos\vartheta. \qquad (4a)$$

(*b*) We may take into account the reflection of the incident wave by the conducting screen and put u_0 equal to incident plus reflected wave. Then we have in the hole $u_0=0$ while the value of $\partial u_0/\partial x'$ is doubled.

Now consider the expression (4) on the screen ($x=0$). Then it follows immediately from (3a) that $\partial\varphi/\partial x=0$ so that the second term in (4) satisfied the boundary condition (2). The first term in (4), however, does not vanish by any means. Unfortunately, it is just this first term which remains present when assumption (*b*) is made. Therefore, for either of the assumptions (*a*) and (*b*), the boundary condition (2) is violated on the "right-hand" (back) side of the screen. This would only be different if we could represent the field by the second term in (4) alone.

The vectorial theory has the same defects. If there are no currents or charges in the half-space $x>0$, the electric field at a point r in that half-space is given by Eq. (19), p. 466, in Stratton's book, which is

$$E(r) = \frac{1}{4\pi}\int d\sigma[ikn\times H(r')\varphi$$

$$-(n\times E(r'))\times\text{grad}\,\varphi - n\cdot E(r')\,\text{grad}\,\varphi]. \qquad (5)$$

In this n is a unit vector in the direction of the inward normal to the surface, i.e., in our case the positive x direction. The first integral in Stratton's formula is omitted because there are assumed to be no charges or currents. The notation is slightly changed (1) because we use Gaussian units so that Stratton's $\omega\mu$ is replaced by k, (2) because we use the gradient with respect to the coordinates of the field point r (point at which the field is to be calculated) rather than of the source point r' (point on the surface where the field is given), (3) because we have interchanged the notations r and r', and (4) because we have reversed the sign of n.

If the integral is extended over the entire surface $x'=0$, Eq. (5) is of course correct. However, according to Kirchhoff's method, the integral is only extended over the hole, and for E and H we insert the field of the wave incident from the left, i.e., E_0, H_0. These quantities may be again considered as constant over the hole which permits integration as in Eq. (4).

Now let us again consider the electric field at a point r on the screen. Then grad φ will be a vector in the plane of the screen, and the same is true of $n\times H$ and of $n\times E$. Therefore the first and last term in (5) give tangential components of E, thus violating the boundary condition $E_{\tan}=0$. Only the second term satisfies the boundary condition, giving an electric field normal to the surface. Again, this second term is the only one which drops out if we include the reflected wave in calculating $E(r')$ and $H(r')$ in the hole.

3. MATHEMATICAL FORMULATION

Let H_0, E_0 be the field on the left-hand side of the screen *if there is no hole*. This field fulfills the boundary condition for $x=0$ (plane of the screen)

$$E_{0\,\tan}=0 \qquad (6)$$

which may also be written in the form

$$n\times E_0=0. \qquad (6a)$$

This makes automatically

$$H_0\cdot n=0. \qquad (6b)$$

$H_{0\,\tan}$, and E_{0n}, are in general different from zero. On the right-hand side of the screen, the field in zero approximation vanishes identically. Then the zero approximation field satisfies the boundary conditions everywhere on the screen but not in the hole: In the hole, H and the normal component of E are discontinuous.

We write now the actual field

$$H=H_0+H_1 \quad \text{for } x<0, \qquad H=H_2 \quad \text{for } x>0, \tag{7}$$

and similarly for the electric field. Then we have the boundary conditions:

$$E_{1\,\tan}=E_{2\,\tan} \quad \text{in the hole,} \tag{8}$$

$$E_{1\,\tan}=E_{2\,\tan}=0 \quad \text{for } x=0 \text{ outside the hole,} \tag{8a}$$

$$H_{2\,\tan}-H_{1\,\tan}=H_{0\,\tan} \quad \text{in the hole.} \tag{8b}$$

The boundary conditions for the normal components are automatically fulfilled if those for the tangential components are satisfied.

It is easily seen that all boundary conditions for E_1 and H_1 are satisfied if E_2 satisfies the boundary condition (8a) and if we put, for any $x>0$, and any y and z:

$$E_{1y}(-x, y, z)=E_{2y}(x, y, z), \tag{9}$$

$$H_{1y}(-x, y, z)=-H_{2y}(x, y, z), \tag{9a}$$

and correspondingly for the z components, Maxwell's equations are consistent with (9), (9a) and will further make

$$H_{1z}(-x, y, z)=H_{2z}(x, y, z), \tag{9b}$$

$$E_{1z}(-x, y, z)=-E_{2z}(x, y, z). \tag{9c}$$

Inserting (9a) into (8b) we find that H_2 must satisfy the boundary condition

$$H_{2\,\tan}=\tfrac{1}{2}H_{0\,\tan} \quad \text{in the hole.} \tag{10}$$

Likewise, from (9c) we get the similar condition

$$E_{2z}=\tfrac{1}{2}E_{0,z} \quad \text{in the hole.} \tag{10a}$$

H_0 and $E_{0,z}$ may be considered as known.

The problem is then to calculate the field E_2, H_2 subject to the boundary conditions (8a), (10), and (10a). These conditions are valid irrespective of size and shape of the hole. However, in the following, we shall assume the hole to be small compared with the wave-length so that H_{0y}, H_{0z}, and E_{0z} may be considered as constant over the hole, and we shall take the shape of the hole to be circular, of radius a.

4. SOLUTION

As we have seen in Section 2, only one of the terms in (5) leads to an acceptable solution, *viz.*, the second. Such a term would be produced by a distribution of "magnetic currents" over the hole (cf. Stratton[1]). We shall therefore *assume a distribution of magnetic currents in the plane of the hole*, but instead of assuming the *current density* to be proportional to $n\times E$, we shall determine it so as to satisfy the boundary conditions (10), (10a).

The magnetic current density J^* and charge density ρ^* can be introduced into Maxwell's equations in the same way as the electric charge and current, *viz.*,

$$\operatorname{div} H=4\pi\rho^*, \tag{11}$$

$$\operatorname{curl} E+\frac{1}{c}\frac{\partial H}{\partial t}=-4\pi J^*. \tag{11a}$$

These equations are identical with Stratton's Eqs. I and III, reference 1, p. 464, except for the units used: We are using Gaussian, non-rational units, and we measure ρ^* in "magnetostatic" units, J^* in "magnetoelectric" units. The continuity equation corresponding to (11), (11a) is

$$\operatorname{div} J^*+\frac{1}{c}\frac{\partial \rho^*}{\partial t}=0. \tag{12}$$

We shall assume the time dependence of all quantities to be as

$$e^{-i\omega t} \tag{12a}$$

so that (3a) represents an outgoing spherical wave. Then (12) becomes

$$\operatorname{div} J^*=ik\rho^* \tag{12b}$$

with

$$k=\omega/c. \tag{12c}$$

It need hardly be pointed out that J^* and ρ^* have no physical meaning.

We shall not use magnetic volume currents and charges but only magnetic surface currents (density K) and surface charges (density η); η corresponds to a discontinuity of H_n at the surface [cf. (11)] and K to a discontinuity of $E_{\tan}$ [cf. (11a)]. The quantities K and η satisfy a continuity equation exactly like (12b), *viz.*,

$$\operatorname{div} K=ik\eta. \tag{13}$$

The electric and magnetic field can be expressed in terms of K and η most conveniently

with the help of a scalar and vector potential, *viz.*,

$$E=\text{curl } F, \tag{14}$$

$$H=-\frac{1}{c}\frac{\partial F}{\partial t}-\text{grad } \psi. \tag{14a}$$

Equations (14) and (14a) automatically satisfy the Maxwell equations

$$\text{curl } H=\frac{1}{c}\frac{\partial E}{\partial t}, \tag{15}$$

$$\text{div } E=0. \tag{15a}$$

From Eqs. (11), (11a) we find ψ and F in exact analogy to the electric case. If only surface charges are present, we obtain

$$F(r)=\int K(r')\varphi(|r-r'|)dr', \tag{16}$$

$$\psi(r)=\int \eta(r')\varphi(|r-r'|)dr'. \tag{16a}$$

Inserting these results, and the time dependence (12a), into (14a) we find the magnetic and electric fields explicitly [cf. Stratton, reference 1, p. 466, Eqs. (19) and (23)]

$$E(r)=\int K(r')\times\text{grad } \varphi\, d\sigma \tag{17}$$

and [Stratton, reference 1, Eqs. (20) and (23)]

$$H(r)=\int (ikK(r')\varphi-\eta(r')\text{ grad }\varphi)d\sigma. \tag{18}$$

The gradients are taken with respect to the coordinates of the field point r (cf. Section 2 for differences between our and Stratton's notation); the integral goes over the area of the hole.

It is seen immediately that (17) satisfied the boundary condition (8a), *viz.*, $E_{\tan}$ vanishes on the screen everywhere outside the hole. In the hole, $E_{\tan}$ is of course not zero but directly related to K [cf. Section 7, Eq. (61)]. However, the problem remains to determine K in such a way as to satisfy the boundary conditions for the magnetic field in the hole, *viz.*, (10), (10a), and this is, of course, a more difficult task.

a. Determination of η from $H_{\tan}$

In the solution of our problem, we are greatly helped by the fact that the hole is small. Then the retardation may be neglected, and H may be considered essentially as the magnetostatic field corresponding to the charge distribution η. The first term in (14a) or (18) is small; indeed, from (13) we find that K is of order ηka where a is the radius of the hole; moreover, grad φ is of the order φ/a so that the first term of (18) is of the order of $(ka)^2$ times the second. Since the hole is assumed to be small compared with the wave-length ($ka\ll 1$), we shall neglect the first term of (18) altogether. Then (14a) reduces to

$$H=-\text{grad } \psi, \tag{19}$$

and since we also neglect the retardation in φ, (16a) reduces to

$$\psi(r)=\int \eta(r')\frac{dr'}{|r-r'|}. \tag{20}$$

In our approximation we may assume $H_{\tan}$ to be constant over the hole and equal to $\frac{1}{2}H_0$ [cf. (10)]. Then we have from (19)

$$\psi=-\tfrac{1}{2}H_0\cdot r. \tag{21}$$

The potential problem given by (20), (21) is fairly well known from electrostatics;[3] we seek a two-dimensional charge distribution which gives a constant field, $\frac{1}{2}H_0$, inside the region occupied by the charges. It is known that a constant inside field *is* produced by a uniform distribution of dipoles in an ellipsoid, the dipoles having (in simple cases) the same direction as the field. If we now assume the x axis of the ellipsoid $2h$ to be very small, the ellipsoidal charge distribution will be equivalent to a surface charge distribution. The cross section of the ellipsoid in the YZ plane should, of course, be taken equal to the hole so that we obtain a rotational ellipsoid with semi-axes, a, a, h. The surface density of dipoles is proportional to the ordinate of the ellipsoid, i.e., proportional to

$$\mu=(a^2-r'^2)^{\frac{1}{2}}. \tag{21a}$$

[3] The following argument was taken from H. Hertz's solution of the elastic problem of the contact between two spheres. See H. Hertz, Crelle's J. 92 (1881) or A. E. H. Love, *Theory of Elasticity*.

The surface *charge* density is then proportional to

$$H_0 \cdot \text{grad}\, \mu = -\frac{H_0 \cdot r'}{(a^2 - r'^2)^{\frac{1}{2}}}. \tag{22}$$

It would also be possible to determine the coefficient of (22) from electrostatics.

We shall, however, determine this coefficient by direct calculation, thus at the same time verifying the solution (22). We put therefore

$$\eta = -C\frac{H_0 \cdot r'}{(a^2 - r'^2)^{\frac{1}{2}}}. \tag{23}$$

In the integral (20), we introduce the coordinates $\rho = |r - r'|$ and β, the angle between the vectors $r' - r$ and r (cf. Fig. 2). Then if α is the angle between r and H_0, we have

$$H_0 \cdot r' = H_0 \cdot r + H_0 \rho \cos(\alpha - \beta). \tag{23a}$$

The integral Eq. (20) becomes, when we insert (21) and (23):

$$C\int_0^{2\pi} d\beta \int_0^{\rho(\beta)} \frac{d\rho}{(a^2 - r'^2)^{\frac{1}{2}}} \times [H_0 \cdot r + H_0 \rho \cos(\alpha - \beta)] = \tfrac{1}{2} H_0 \cdot r. \tag{24}$$

The factor ρ from the element of area cancels the factor $\varphi = 1/|r - r'|$ in (20). It is more convenient to integrate in (24) along a whole chord such as RS in Fig. 2, instead of separately over the parts RP and PS. Then the integral over β goes only from 0 to π, and assumes positive (on PS) as well as negative (on RP) values.

Now we find from Fig. 2,

$$a^2 - r'^2 = a^2 - OQ^2 - QT^2 = s^2 - x^2, \tag{25}$$

where $s = QS$ is half the length of the chord and

$$x = QT = \rho + QP = \rho + r\cos\beta. \tag{25a}$$

x goes from $-s$ to $+s$, and we have

$$\int_{-s}^{s} \frac{dx}{(a^2 - r'^2)^{\frac{1}{2}}} = \pi, \tag{26}$$

$$\int_{-s}^{s} \frac{\rho\, dx}{(a^2 - r'^2)^{\frac{1}{2}}} = -\pi r \cos\beta. \tag{26a}$$

Then (24) becomes

$$C\pi \int_0^{\pi} d\beta (H_0 \cdot r - H_0 r \cos\beta) \cos(\alpha - \beta) = \tfrac{1}{2} H_0 \cdot r. \tag{27}$$

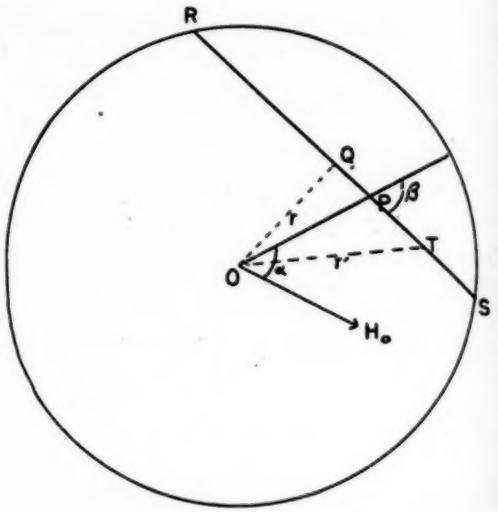

FIG. 2. Illustrating the integration, Eq. (24) to (26). O is the center of the hole, $r = 0$, P is the field point r, RS any chord through P, T any point r' on this chord.

Integration over β gives

$$C\pi \cdot \pi (H_0 \cdot r - \tfrac{1}{2} H_0 r \cos\alpha) = \tfrac{1}{2} C\pi^2 H_0 \cdot r \tag{27a}$$

so that

$$C = 1/\pi^2. \tag{27b}$$

The assumption (23) therefore actually solves the integral equation, and the magnetic charge density is

$$\eta = -\frac{1}{\pi^2 (a^2 - r'^2)^{\frac{1}{2}}} H_0 \cdot r'. \tag{28}$$

From the charge density we can obtain the current density K, using (13). We find

$$K = \frac{ik}{\pi^2} (a^2 - r'^2)^{\frac{1}{2}} H_0. \tag{29}$$

Remembering that H_0 has vanishing normal component everywhere on the plane $x = 0$, we see that K is entirely in the plane of the hole, as it should be.

b. Boundary Condition for E_n

The correct normal component of E_0, *viz.*, $E_{2n} = \tfrac{1}{2} E_{0n}$, must also be obtained by a suitable distribution of magnetic currents K, from (17). It can easily be seen that the distribution (29) does not contribute appreciably to (17). Expres-

sion (29) is of the order kaH_0, grad φ is of order $1/a^2$, and the integral (17) goes over an area of order a^2. Therefore the contribution of (29) to (17) is at most of order kaH_0 while actually the normal component of E_0 is of the same order as H_0. [In reality, the spatial dependence introduced by the grad operator in (17) makes the contribution of (29) of the order $(ka)^2H_0$ only.]

Therefore we must obtain an additional magnetic current distribution K_E to fit the boundary condition for E_n. However, we must take care that the new current distribution does not give rise to an additional magnetic charge density η which might destroy the agreement obtained above for the magnetic field. This is most easily achieved by letting the magnetic current lines be closed circles. Mathematically we must have

$$\operatorname{div} K_E = 0.$$

For the actual evaluation it is most convenient to use (14), (16). The given electric field E is normal to the surface and had the value $\frac{1}{2}E_{0,z}$ which may be considered constant over the hole. Therefore (14) gives

$$F = \tfrac{1}{4}E_0 \times r. \tag{30}$$

The integral equation for the components of K is the same as for η in Section 2, e.g.,

$$\int K_y \frac{dy'dz'}{|r-r'|} = -F_y = \tfrac{1}{4}E_{0,z}z, \tag{31}$$

which corresponds to (20), (21). Therefore we have, in analogy to (28),

$$K_y = \frac{1}{2\pi^2(a^2-r'^2)^{\frac{1}{2}}}E_{0,z}z' \tag{31a}$$

and altogether for the current giving the required value of E.

$$K_E = \frac{1}{2\pi^2(a^2-r'^2)^{\frac{1}{2}}}r' \times E_0. \tag{32}$$

It is easily seen that this solution satisfies the condition (30a). Moreover, since E_0 is normal to the plane, the vector K_E is entirely in the plane of the hole as it should be. The contribution of K_E to the first term in (18) can be shown to be of order $(ka)^2E_{0n}$ which is negligible.

c. Final Formula

The total magnetic current and charge density in the hole is now

$$K = K_H + K_E = \frac{1}{\pi^2}\left(ik(a^2-r'^2)^{\frac{1}{2}}H_0 + \frac{1}{2(a^2-r'^2)^{\frac{1}{2}}}r' \times E_0 \right), \tag{33}$$

$$\eta = -\frac{1}{\pi^2(a^2-r'^2)^{\frac{1}{2}}}r' \cdot H_0. \tag{33a}$$

5. CALCULATION OF THE DIFFRACTED FIELD

The diffracted field E, H for $x>0$ can be calculated by inserting (33), (33a) into (17) and (18). According to its construction, this field will satisfy Maxwell's equations and all boundary conditions.

We shall carry out the integration at large distance from the hole, $kr \gg 1$. Let κ be a unit vector in the direction of r, i.e., in the direction of propagation of the diffracted wave. Then we have

$$\operatorname{grad} \varphi = ik\kappa\varphi. \tag{34}$$

In some of the integrals we may replace φ by its value for $r'=0$, *viz.*,

$$\varphi_0 = e^{ikr}/r. \tag{35}$$

In other cases, it is necessary to expand φ in powers of r' and keep the linear term; then we obtain

$$\varphi = \varphi_0(1 - ik\kappa \cdot r'). \tag{35a}$$

a. Evaluation of the Electric Field

According to (17) and (34), the electric field is

$$E(r) = -ik\kappa \times \int K\varphi dy'dz'. \tag{36}$$

It is convenient to calculate separately the contributions of K_H and K_E [cf. (33)]. We have

$$\int K_H \varphi dy'dz' = \frac{ik}{\pi^2}H_0\varphi_0 \int_0^a (a^2-r'^2)^{\frac{1}{2}} 2\pi r' dr' = \frac{2i}{3\pi}ka^3H_0\varphi_0. \tag{37}$$

The contribution of K_E is slightly more difficult to calculate because K_E contains the factor r' and therefore φ must be expanded in the form (35a). The first term, φ_0, does not give any contribution so that we obtain

$$\int K_E \varphi dy' dz' = -\frac{ik\varphi_0}{2\pi^2} \times \int_0^{2\pi} d\beta \int_0^a \frac{r'dr'}{(a^2-r'^2)^{\frac{1}{2}}} \kappa \cdot r'\, r' \times E_0, \quad (37a)$$

where r' and β are polar coordinates in the plane of the hole. Remembering that E_0 is in the x direction (normal to the hole), we find

$$\int d\beta \kappa \cdot r' r' \times E_0 = -\pi r'^2 E_0 \times \kappa \quad (37b)$$

and

$$\int K_E \varphi dy' dz' = \frac{i}{3\pi} k a^3 E_0 \times \kappa \varphi_0. \quad (38)$$

Inserting (37) and (38) into (36) we find

$$E = \frac{1}{3\pi} k^2 a^3 \varphi_0 \kappa \times (2H_0 + E_0 \times \kappa). \quad (39)$$

We see that E is always perpendicular to κ, i.e., the waves are transverse. Moreover, for a point r on the conducting screen, κ lies in the plane of the screen and H_0 and $\kappa \times E_0$ do likewise. Therefore E is normal to the screen thus satisfying the boundary conditions.

b. Magnetic Field

The first term of (18) follows directly from (37), (38):

$$H^{(1)} = ik \int K \varphi d\sigma = -\frac{1}{3\pi} k^2 a^3 \varphi_0 (2H_0 + E_0 \times \kappa). \quad (40)$$

This field is not transversal; only the second term ($E_0 \times \kappa$) has this property. From (39) and (40) we see that

$$E = -\kappa \times H^{(1)}. \quad (40a)$$

The second term of (18) gives

$$H^{(2)} = -\int \eta \operatorname{grad} \varphi d\sigma = -k^2 \kappa \varphi_0 \int \eta \kappa \cdot r' d\sigma. \quad (41)$$

Inserting (33a) we get

$$H^{(2)} = \frac{k^2}{\pi^2} \kappa \varphi_0 \int_0^{2\pi} d\beta \int_0^a \frac{r'dr'}{(a^2-r'^2)^{\frac{1}{2}}} \kappa \cdot r' H_0 \cdot r'. \quad (41a)$$

Now

$$\int d\beta \kappa \cdot r' H_0 \cdot r' = \pi r'^2 H_0 \cdot \kappa \quad (41b)$$

and therefore

$$H^{(2)} = \frac{2}{3\pi} k^2 a^3 \varphi_0 \kappa\, H_0 \cdot \kappa. \quad (42)$$

Remembering the vector identity,

$$H_0 - H_0 \cdot \kappa\, \kappa = \kappa \times (H_0 \times \kappa), \quad (42a)$$

we find the entire magnetic field, (40)+(42), becomes

$$H = H^{(1)} + H^{(2)} = -\frac{1}{3\pi} k^2 a^3 \varphi_0 \kappa \times (2H_0 \times \kappa - E_0). \quad (43)$$

This expression is a transverse wave as it should be. Thus the field $H^{(2)}$ serves to satisfy the transversality condition. In fact, since (42) is entirely longitudinal (in direction of κ), $H^{(2)}$ *serves only to eliminate the longitudinal component of* $H^{(1)}$ *and leaves the transverse component unchanged.* This result will be important in the application to the cavity problem (Section 9).

Comparing (43) and (39) we see that

$$H = \kappa \times E \quad (44)$$

$$E = -\kappa \times H \quad (44a)$$

as is required by Maxwell's equations for plane waves.

c. Total Radiation

Poynting vector of the diffracted field is

$$S = \frac{c}{4\pi} E \times H = \frac{c}{36\pi^3} \frac{k^4 a^6}{r^2} \kappa (2\kappa \times H_0 - \kappa \times \kappa \times E_0)^2. \quad (45)$$

Now let θ denote the angle between the propagation vector κ and the vector n normal to the screen (x direction) and α the azimuth of κ, i.e., the angle between the plane of κ and n and the plane of H_0 and n. Then the intensity of radiation

in the direction κ is, per unit solid angle,

$$r^2|S| = \frac{c}{36\pi^3}k^4a^6[4H_0^2\cos^2\theta\cos^2\alpha + (\sin\theta E_{0,x} - 2H_0\sin\alpha)^2]. \quad (46)$$

The radiation is thus not symmetrical about any axis. The total radiation in all directions is

$$S_{tot} = \int_0^{\pi/2}\sin\theta d\theta\int_0^{2\pi}d\alpha r^2|S| = \frac{c}{27\pi^2}k^2a^6(4H_0^2+E_0^2). \quad (47)$$

The expressions E_0^2 and H_0^2 on the right-hand side denote the time averages.

6. DISCUSSION OF THE RESULT

a. Comparison with the Kirchhoff Solution

The result for the diffracted field is entirely different from that of the Kirchhoff theory. Quite obviously, the polarization of the radiation is different because our solution satisfies the boundary condition while Kirchhoff's does not. Also the angular distribution of the total radiation is different (cf. Subsection c, below). However, the most striking difference is the absolute value of the field. As is easily seen from expression (5), the Kirchhoff solution gives magnetic and electric fields of the order

$$H_K \sim ka^2H_0.$$

Our solution gives instead [cf. (39), (43)]

$$H_B \sim k^2a^3H_0$$

which differs from H_K by a factor ka. In addition, our solution has a smaller numerical factor. Therefore, for small holes, the radiation transmitted through the hole is very much smaller than Kirchhoff's theory would indicate. The fields E and H are reduced roughly in the ratio a/λ where λ is the wave-length of the radiation. The radiation intensity is therefore reduced by a factor of the order $(a/\lambda)^2$.

b. Representation by Electric and Magnetic Dipoles

Turning now to our solution itself, we note that the field may be considered as owing to an electric and a magnetic dipole. The field of a magnetic dipole of moment M at the origin is, according to usual electromagnetic formulas:

$$H = k^2\varphi_{0\kappa}\times(M\times\kappa), \quad (48)$$

$$E = -k^2\varphi_{0\kappa}\times M. \quad (48a)$$

Similarly, the field of an electric dipole of moment P is

$$E = k^2\varphi_{0\kappa}\times(P\times\kappa) \quad (49)$$

$$H = k^2\varphi_{0\kappa}\times P. \quad (49a)$$

Comparison with (39) or (43) shows that the hole is equivalent to a magnetic dipole

$$M = -\frac{2}{3\pi}a^3H_0 \quad (50)$$

and an electric dipole

$$P = -\frac{1}{3\pi}a^3E_0. \quad (50a)$$

Each of the dipoles is antiparallel to the respective field in the hole as might be expected. Accordingly, the magnetic dipole is in the plane of the hole and the electric one in the direction of the normal. These directions are just those required to fulfill the boundary conditions for the electric field on the screen. The magnetic moment (50) can easily be calculated from the expression (33a) for the magnetic charge density; we have

$$M = \int\eta r'dr' = -\frac{1}{\pi^2}\int r'\cdot H_0\frac{r'}{(a^2-r'^2)^{\frac{1}{2}}}r'dr'd\beta = -\frac{1}{\pi}H_0\int r'dr'\frac{r'^2}{(a^2-r'^2)^{\frac{1}{2}}} = -\frac{2}{3\pi}H_0a^3. \quad (51)$$

It will be remembered that the solution in Section 4a was actually obtained by assuming a distribution of magnetic dipoles in a flat ellipsoid. As is well known from electrostatics, the inner field produced by such a distribution is proportional to the polarization per unit volume (for any given shape of the ellipsoid); the total moment is therefore proportional to H_0 times the volume which in turn is proportional to a^3.

Dimensional considerations, in combination with the fact that the relation between M and H_0 is a purely magnetostatic problem, also show that the moment M must be proportional to H_0a^3.

The different factor (1 instead of 2) in the electric moment (50a) is no doubt owing to the fact that the electric moment is in the direction of the small axis of the ellipsoid while the magnetic moment is in the direction of one of the long axes.

The order of magnitude of M and P is in general the same although there are cases for which one of them vanishes. This is frequently the case for the electric moment P: It is only necessary to assume incident plane waves with the electric field perpendicular to the plane of incidence. Then the electric field E_0 will have no component perpendicular to the screen, and only H_0 will be different from zero. The radiation intensity is then proportional to $\sin^2\chi$ where χ is the angle between the direction of H_0 and the direction of propagation of the diffracted waves κ. This radiation is large in the normal direction, and also on the screen in the direction perpendicular to H_0, i.e., in the direction of the electric vector of the incident wave.

To make H_0 vanish and $E_0\neq 0$, it is necessary to assume standing waves on the "left-hand side" of the screen. For example, in a rectangular cavity with the sides $L_1L_2L_3$, the relevant components of the electric and magnetic fields are given by

$$E_x=A_x\cos\frac{\pi m_1x}{L_1}\sin\frac{\pi m_2y}{L_2}\sin\frac{\pi m_3z}{L_3}, \tag{52}$$

$$H_y=B_y\cos\frac{\pi m_1x}{L_1}\sin\frac{\pi m_2y}{L_2}\cos\frac{\pi m_3z}{L_3}, \tag{52a}$$

$$H_z=B_z\cos\frac{\pi m_1x}{L_1}\cos\frac{\pi m_2y}{L_2}\sin\frac{\pi m_3z}{L_3}, \tag{52b}$$

where A_x, B_y, and B_z are constants and m_1, m_2, m_3 integers. Then if we place the hole on the wall $x=0$ at a point where

$$\frac{m_2y}{L_2}=n_2+\tfrac{1}{2} \quad\text{and}\quad \frac{m_3z}{L_3}=n_3+\tfrac{1}{2} \tag{52c}$$

(n_2 and n_3 integers), we have $H_{0y}=H_{0z}=0$ and $E_{0z}\neq 0$. Then the radiated intensity [cf. (46)] is proportional to $\sin^2\theta$ where θ is the angle between the normal to the screen (direction of E_0) and the direction of propagation. This intensity is a maximum on the screen and zero in the normal direction, quite different from the previous case.

c. Diffraction of a Plane Wave

It is of some interest to discuss the case of an incident plane wave. Let H_i, E_i, κ_i be, respectively, the magnetic field, the electric field, and a unit vector in the direction of propagation for the incident wave. Analogously to (44), (44a) we have

$$H_i=\kappa_i\times E_i, \tag{53}$$

$$E_i=-\kappa_i\times H_i, \tag{53a}$$

and the Poynting vector is

$$S_i=\kappa_i\frac{c}{4\pi}E_i^2=\kappa_i\frac{c}{4\pi}H_i^2. \tag{54}$$

We shall denote by ϑ the angle of incidence, i.e., between κ_i and n (the vector n is in the x direction); then $0<\vartheta<\pi/2$. We have to distinguish two cases.

α. Electric Field of Incident Wave Perpendicular to Plane of Incidence

For this case we have E_i perpendicular to the plane of κ_i and n. Then E_i has no normal component, and the tangential component of H_i is

$$H_{i\,\mathrm{tan}}=H_i\cos\vartheta. \tag{55}$$

This tangential component is doubled upon reflection so that $H_0=2H_{i\,\mathrm{tan}}$; it lies in the plane of incidence. If, then, β is the angle between the plane of κ and n and the plane of κ_i and n, we have for the diffracted radiation intensity per unit solid angle

$$r^2|S|=\frac{4c}{9\pi^3}k^4a^6H_i^2\cos^2\vartheta(1-\sin^2\theta\cos^2\beta)$$

$$=\frac{16}{9\pi^2}k^4a^6S_i\cos^2\vartheta(1-\sin^2\theta\cos^2\beta). \tag{56}$$

The "diffraction cross section" of the hole for this polarization is then

$$A_\perp = \frac{\int |S| r^2 \sin\theta d\theta d\beta}{S_i} = \frac{64}{27\pi} k^4 a^6 \cos^2\vartheta. \quad (57)$$

β. Electric Field in Plane of Incidence

If we choose the plane of incidence as XY plane, then κ_i has the components $(\cos\vartheta, \sin\vartheta, 0)$ and E_i has the components $E_i(-\sin\vartheta, +\cos\vartheta, 0)$. According to (53), H_i has then the components $E_i(0, 0, 1)$. The components of κ are $(\cos\theta, \sin\theta\cos\beta, \sin\theta\sin\beta)$. Taking into account the reflection and the fact that only the normal component of E_i is effective, we get

$$E_0 = -2E_i(\sin\vartheta, 0, 0),$$
$$H_0 = 2E_i(0, 0, 1), \quad (57a)$$

and [cf. (45)]

$$\begin{aligned}(2\kappa\times H_0 - \kappa\times\kappa\times E_0)^2 &= 4(\kappa\times H_0)^2 + (\kappa\times E_0)^2 + 4\kappa\cdot(H_0\times E_0) \\ &= 4E_i^2[4(1-\sin^2\theta\sin^2\beta) + \sin^2\theta\sin^2\vartheta \\ &\qquad - 4\sin\theta\sin\vartheta\cos\beta]. \quad (57b)\end{aligned}$$

Inserting into (45) we find for the differential cross section

$$\begin{aligned}\frac{dA}{\sin\theta d\theta d\beta} &= \frac{r^2|S|}{S_i} \\ &= \frac{16}{9\pi^2} k^4 a^6[\cos^2\theta + \sin^2\theta(\cos^2\beta \\ &\qquad + \tfrac{1}{4}\sin^2\vartheta) - \sin\theta\sin\vartheta\cos\beta] \quad (58)\end{aligned}$$

and for the total cross section

$$A_\parallel = \frac{64}{27\pi} k^4 a^6 (1 + \tfrac{1}{4}\sin^2\vartheta). \quad (59)$$

The most remarkable feature of the angular distribution (58) is the dependence on the azimuth β: For a given θ, the radiation is smaller for $\beta = 0$ (i.e., in the direction closest to that of the incident wave), than in the "reflected" direction $\beta = \pi$. We suspected an error in sign to be the cause of this result, but we have found none. Perhaps the most convincing check on the sign is provided by Eqs. (48) to (50a). The result mentioned is of course exactly the opposite of that expected from any elementary considerations based on the Huygens principle. For any value of θ, the radiation is a minimum in the azimuth given by

$$\cos\beta_0 = \tfrac{1}{2}\sin\vartheta/\sin\theta. \quad (58a)$$

In the plane of the screen, the radiation is zero for $\beta = \beta_0$, which in this case lies between 60° and 90°.

The total cross section (59) is greater than that for the α-direction of polarization, (57), except for normal incidence when the two expressions are equal. For unpolarized light,

$$A' = \frac{64}{27\pi} k^4 a^6 (1 - \tfrac{3}{8}\sin^2\vartheta). \quad (59a)$$

All cross sections are proportional to λ^{-4} as in Rayleigh's theory of the scattering by small objects. Also the proportionality with a^6 is the same as in the scattering theory. The cross section is of the same order as the scattering cross section of a dielectric sphere or disk of radius a and dielectric constant of the order of 2.

d. Comparison with Sommerfeld's Solution

One characteristic feature of our solution is that the current and charge distribution K, η determine the field only in the half-space on the right of the screen. The perturbation of the field on the left, H_1 and E_1 [cf. (7), (8)] would be obtained by a current and charge distribution $-K$, $-\eta$. This reversal of sign is unusual in electrodynamic theory. But a somewhat similar procedure had to be used by Sommerfeld in the wedge problem. In order to take into account that the field is entirely different on the two sides of the screen, Sommerfeld put a "branch plane" in the screen and obtained the solution of the problem in a Riemann space. This is the mathematical expression for the fact that the field is not given by the same integral representation on the two sides of the screen.

7. FIELD IN AND NEAR THE HOLE

a. Inside the Hole

From the conditions (10), (10a), (7), (9a), (9c), we see immediately that in the hole

$$H_{\tan}=\tfrac{1}{2}H_{0\,\tan}, \tag{60}$$

$$E_n=\tfrac{1}{2}E_{0,n}. \tag{60a}$$

$H_{\tan}$ and E_n are thus halfway between their values in the "unperturbed" fields on the right and on the left side of the screen. On the other hand, those field components which would be zero if the hole were absent, i.e., the normal component of H and the tangential component of E, are determined directly by η and K. Since $4\pi\eta$ measures the discontinuity[4] of H_n, we have, at a point just to the right of the hole,

$$H_z(r)=2\pi\eta=-\frac{2}{\pi}\frac{H_0\cdot r}{(a^2-r^2)^{\frac{1}{2}}}. \tag{60b}$$

In the hole, therefore, the normal component of H is of the same order of magnitude as the tangential component. This does not in any way invalidate our calculations because we have not used a perturbation method in which H_z was considered small compared with $H_{\tan}$, but we have satisfied the boundary conditions exactly.

In contrast to $H_{\tan}$ which was assumed to be constant, H_z varies rapidly over the surface of the hole. Near the boundary $r=a$, H_z becomes infinite as $1/(a-r)^{\frac{1}{2}}$. An integrable infinity of this type occurs frequently at edges. It occurs, e.g., in Summerfeld's solution of the diffraction by a half-plane, in the neighborhood of the diffracting edge.

The tangential component of the electric field is given by [cf. Stratton, p. 467, Eq. (23)]

$$E_{\tan}=2\pi n\times K, \tag{61}$$

where n is in the positive x direction (Stratton's n is in our negative x direction). Inserting (33) and remembering that in the hole $r\cdot n=0$, we have

$$E_{\tan}=\frac{1}{\pi}\frac{\mathbf{r}}{(a^2-r^2)^{\frac{1}{2}}}E_{0,z}+\frac{2}{\pi}ik(a^2-r^2)^{\frac{1}{2}}n\times H_0. \tag{61a}$$

The second term is in general negligible compared with the first since H_0 and E_{0z} are usually of the same order of magnitude. The first term is again of the same order as E_{0z} [cf. (60b)] and is singular near the edge of the hole. The tangential electric field is directed radially outwards from the center of the hole.

b. E_{norm} and $H_{\tan}$ near the Hole

Outside the hole, $E_{\tan}$ and H_n are of course zero. The other field components have singularities near the edge, behaving as $1/(r^2-a^2)^{\frac{1}{2}}$. At larger distance from the hole, $H_{\tan}$ and E_n behave as H_0a^3/r^3, up to distances of the order λ. For still larger distances, the solutions of Section 5 hold. The field in the neighborhood of the hole has not yet been calculated in detail. It is clear, however, that $H_{\tan}$ and E_n are very far from zero on the screen in contrast to the assumption made in Kirchhoff's theory.

The tangential component of H is especially important because it determines the dissipation of energy (joule heat) in the screen if the latter has finite conductivity. The energy dissipation per unit time is

$$P=\tfrac{1}{4}\nu D\int H_{\tan}^2 dA, \tag{62}$$

where $\nu=\omega/2\pi$ is the frequency,

$$D=\frac{1}{2\pi(\sigma\nu)^{\frac{1}{2}}} \tag{62a}$$

is the thickness of the current carrying sheath in the conductor, σ is the conductivity in electromagnetic units ($=10^{-9}$ times conductivity in mho), and the magnetic permeability has been assumed to be 1. The integral in (62) goes over the whole surface of the conductor. Since $H_{\tan}$ behaves as $1/(r^2-a^2)^{\frac{1}{2}}$ near the edge of the hole, (62) diverges logarithmically; it must be cut off at a value of $r-a$ of the order of D. The power dissipation on the right-hand surface of the screen becomes then of the order

$$P\sim H_0^2\nu Da^2\ln(a/D). \tag{62b}$$

An accurate calculation of this quantity would be useful for the theory of cavities with small holes.

[4] The discontinuity of H_n at $x=0$ is, of course, only mathematical. As explained in Section 6d, the current distribution K, η determines the field only for $x>0$. The physical field H_n is continuous at $x=0$, according to (9b).

8. POSSIBLE EXTENSIONS OF THE THEORY

In its present form, the theory is only applicable to holes small compared with the wave-length. However, it seems possible to extend it to holes comparable in size with the wave-length. In this case, the given fields H_0 and E_0 will contain factors of the type $e^{ik\cdot r'}$. Similarly, the variation of the factor $e^{ik|r-r'|}$ in the Green's function φ must be considered. From an approximate consideration, it seems that the correction terms will be of relative order $(ka)^2$ rather than ka.

It seems certainly possible to obtain solutions in terms of power series in y' and z'. The integral (20), for instance, can be evaluated if η is any odd power of $(a^2-r'^2)^{\frac{1}{2}}$, giving for H_{0y} a power series in r^2. From this type of solution, others can be obtained by differentiation with respect to y and z. We believe that in this way a solution for an arbitrary function $H_{0y}(y, z)$ can be obtained. However, a more elegant solution may perhaps be found in terms of the electric oscillations of an ellipsoid.

A second question is the dependence on the shape of the hole. From the way in which our solution was obtained in Section 4a, it is clear that a solution can also be found for an elliptical hole. The solution for an ellipse of arbitrary eccentricity should give a sufficient idea about the dependence of the diffraction on the shape of the hole. However, the case of a rectangle may also be solvable.

A more difficult question is the extension to screens and holes which are not plane. The case of a small hole in a curved surface will probably still give a result very similar to ours, as long as the surface may be considered plane over a region large compared with the hole. This is usually the case since the radius of curvature of cavities is ordinarily of the order of the wave-lengths. If this condition is fulfilled, the boundary condition on the surface will be violated only in the wave zone where the diffracted field is no longer very large. Perhaps a method of successive approximations will be applicable to this case.

However, a different situation arises when the hole itself is essentially curved. A simple example is a gap in a cylindrical wave guide; in this case, the gap itself is the surface of a short cylinder. Such a problem will presumably require an entirely new solution, but it is likely that at least the same principle will work as in our case, and that the symmetry of the cylinder will be helpful for obtaining the solution.

9. APPLICATION TO THE THEORY OF CAVITIES

Condon[5] has given a most convenient theory of the excitation of cavities by electric current loops. In this theory, the vector potential A is developed in terms of normal modes A_m, thus:

$$A=\sum_m p_m(t)A_m(r). \tag{63}$$

By inserting this expression into Maxwell's equations, a differential equation is obtained for p_m in terms of the current I exciting the cavity, *viz.*:

$$\frac{d^2p_m}{dt^2}+\frac{\omega_m}{Q_m}\frac{dp_m}{dt}+\omega_m{}^2p_m=\frac{4\pi I}{V}\int A_m\cdot ds, \tag{64}$$

where ω_m is the frequency of the mth normal mode, Q_m is the dissipation constant, V is the volume of the cavity, and the integral is taken along the conductor carrying the current I. The current is measured in electromagnetic units.

Since our solution is obtained in terms of magnetic currents and charges, it is convenient to expand the field in a slightly different way, by using a magnetic vector potential F as defined in (14), (14a). The scalar potential ψ will be unimportant because the magnetic field is transverse. We expand:

$$F=\sum_m q_m(t)F_m(r) \tag{65}$$

and have

$$H=\frac{1}{c}\sum_m\frac{dq_m}{dt}F_m(r), \tag{65a}$$

$$E=\sum_m q_m k_m A_m(r) \tag{65b}$$

with

$$A_m=\text{curl } F_m/k_m. \tag{65c}$$

The F_m form an orthonormal set:

$$\int F_l\cdot F_m dV=V\delta_{lm}, \tag{65d}$$

where V is the volume of the cavity. Our q_m and

[5] E. V. Condon, J. App. Phys. 12, 129 (1941).

Condon's p_m stand in the relation of coordinate and momentum.

The solution (65) satisfies automatically the first Maxwell equation (15), (15a) (cf. Section 4).

The F_m's are determined in such a way that also

$$\text{div } F_m = 0. \tag{66}$$

Moreover, the F_m satisfy the wave equation

$$\nabla^2 F_m + k^2 F_m = 0 \tag{66a}$$

with

$$k_m = \omega_m / c. \tag{66b}$$

Because of (15), (66), (66a) the second Maxwell equation

$$\text{curl } E = -\frac{1}{c}\frac{\partial H}{\partial t} \tag{66c}$$

is also fulfilled by the particular solutions F_m.

Now assume a distribution of magnetic currents of current density J^* per unit volume. Then the Maxwell equation (66c) is replaced by (11a). Inserting Eqs. (14), (14a) we obtain

$$\text{curl curl } F + \frac{1}{c^2}\frac{\partial^2 F}{\partial t^2} = -4\pi J^*. \tag{66d}$$

Using the expansion (65), the condition (66), and the wave equation (66a), and expanding the right-hand side of (66d) in terms of eigenfunctions F_m, we obtain

$$\frac{d^2 q_m}{dt^2} + \frac{\omega_m}{Q_m}\frac{dq_m}{dt} + \omega_m^2 q_m = -\frac{4\pi c^2}{V}\int dV J^* \cdot F_m. \tag{67}$$

The second term has been added to take into account the energy dissipation in the walls of the cavity. If we have magnetic surface currents K instead of volume currents J^*, the right-hand side of (67) is replaced by

$$-\frac{4\pi c^2}{V}\int d\sigma K \cdot F_m. \tag{67a}$$

Equations (67), (67a) are obviously the exact analog of Condon's Eq. (64).

It will be noted that (67a) depends only on K, not on η. Indeed, as we have shown in Section 5b, the magnetic charge η serves only to eliminate the longitudinal component of the magnetic field and does not affect the transverse component at all.

Now the normal modes of the cavity all have transverse magnetic fields. Therefore to determine their excitation we need only the magnetic current distribution K, not η. Moreover, our expansion in transverse modes will give us the complete field because the longitudinal field produced by K is canceled by the effect of η.

We may now evaluate (87a) in terms of $H_{0\,\tan}$ and E_{0n}. Inserting (33), we find

$$\int K_H \cdot F_m d\sigma = \frac{ik}{\pi^2} H_0 \cdot F_m(0) \int (a^2 - r'^2)^{\frac{1}{2}} 2\pi r' dr'$$

$$= \frac{2i}{3\pi} k a^3 H_0 \cdot F_m(0), \tag{68}$$

where $F_m(0)$ is the value of the "eigenfunction" F_m at the center of the hole. To evaluate the contribution of K_E, Eq. (33), we must expand F_m in powers of r' and keep the linear term:

$$F_{my} = F_{my}(0) + \frac{\partial F_{my}}{\partial y'} y' + \frac{\partial F_{my}}{\partial z'} z'. \tag{68a}$$

The constant term in (68a) gives no contribution, and we obtain

$$\int K_E \cdot F_m d\sigma = \frac{E_{0z}}{2\pi^2}\int_0^{2\pi} d\beta \int_0^a \frac{r' dr'}{(a^2 - r'^2)^{\frac{1}{2}}}$$

$$\cdot \left[\left(\frac{\partial F_{my}}{\partial y'} y' + \frac{\partial F_{my}}{\partial z'} z'\right) z' - \left(\frac{\partial F_{mz}}{\partial y'} y' + \frac{\partial F_{mz}}{\partial z'} z'\right) y'\right]$$

$$= -\frac{1}{3\pi} a^3 E_{0z}\, \text{curl}_z F_m$$

$$= -\frac{1}{3\pi} k_m a^3 E_{0z} A_{mz}(0). \tag{69}$$

In the last transformation, we have used relation (65c). Inserting (68), (69) into (67) we obtain

$$\frac{d^2 q_m}{dt^2} + \frac{\omega_m}{Q_m}\frac{dq_m}{dt} + \omega_m^2 q_m = \frac{4}{3}\frac{c^2 a^3}{V}[k_m E_{0z} A_{mz}(0) - 2ikH_0 \cdot F_m(0)]. \tag{70}$$

We must now determine H_0 and E_0. The simplest case, perhaps, is that of a cavity communicating with outside space through a small hole. If radiation comes from the outside and if there is no radiation in the cavity, then H_0, E_0 are simply the incident magnetic and electric field at the hole. If there is also a field in the cavity, it is easy to see that in order to satisfy the boundary conditions in the hole, we must choose H_0 and E_0 equal to the *difference* between the outside and the inside field.

A case of some interest is that of a cavity with a field inside, radiating into space through a small hole. Then H_0, E_0 must be replaced by the inside field at the hole, with negative sign. Some caution is necessary since it has been assumed in the derivation in Section 4 that the fields depend on time as $e^{-i\omega t}$. Therefore, if the inside field H is to be obtained from (65a), we must put[6]

$$(1/c)(dq_m/dt) = -ikq_m. \tag{70a}$$

Now let us suppose that the frequency is close to the natural frequency ω_1 of a certain mode, 1, of the cavity. Then we may put in sufficient approximation

$$E_{0x} = -q_1 k_1 A_{1x}, \tag{71}$$

$$H_0 = -\frac{1}{c}\frac{dq_1}{dt}F_1 = ik_1 q_1 F_1 \tag{71a}$$

and obtain

$$\frac{d^2q_m}{dt^2}+\frac{\omega_m}{Q_m}\frac{dq_m}{dt}+\omega_m{}^2 q_m = \frac{4}{3}\frac{a^3}{V}\omega_1 q_1[2\omega F_1(0)\cdot F_m(0) - \omega_m A_{1x}(0)A_{mx}(0)], \tag{72}$$

where the argument 0 denotes the value of the respective function at the center of the hole.

Equation (72) describes the excitation of all normal modes m by one normal mode, 1, through the action of the hole. In the stationary state, i.e., when

$$q_1 = A \sin \omega_1 t \tag{73}$$

the excitation of the mth mode will be given by

$$q_m = \frac{4}{3}A\frac{a^3}{V}\sin \omega_1 t \frac{\omega_1}{\omega_m{}^2-\omega_1{}^2} \times[2\omega_1 F_1(0)\cdot F_m(0) - \omega_m A_{1x}(0)A_{mx}(0)]. \tag{73a}$$

Thus quite similar to the perturbation theory of wave mechanics, the other modes m will be "mixed in" with the main mode on account of the "perturbation" by the small hole. However, it is in general not necessary to use the coefficients q_m because the field distribution near the hole is given more directly by the integrals (17), (18) and is only very little modified by the presence of the rest of the cavity (for given H_0, E_0).

The most interesting result is obtained by applying Eq. (72) to the main mode $m=1$ which gives

$$\frac{d^2q_1}{dt^2}+\frac{\omega_1}{Q_1}\frac{dq_1}{dt}+\omega_1{}^2 q_1 = 2\omega_1{}^2\gamma q_1, \tag{74}$$

where the emission coefficient γ is

$$\gamma = \frac{2a^3}{V}[2F_1{}^2(0) - A_{1x}^2(0)]. \tag{74a}$$

The right-hand side of (74) will cause a shift of frequency from ω_1 to

$$\omega_1' = \omega_1(1-\gamma). \tag{75}$$

The amount and sign of this shift depends on the value of F_1 and A_{1x} at the hole. If A_{1x} is small, the frequency will be decreased; if $F_1 \ll A_{1x}$, there will be an increase of frequency.

Both alternatives may occur in practice: E.g., in a rectangular cavity with electric field in the x direction, the field will be given by (52) with $m_1=0$. Then the electric field will be zero on the XY and the XZ plane; a hole in either of these planes will therefore *lower* the natural frequency. On the YZ plane, at the points determined by (52c), $F_1=0$ and A_{1x} is a maximum; therefore a hole at these positions will increase the frequency. If $m_2=m_3=1$, this is the case in the middle of the YZ plane ($y=\frac{1}{2}L_2$, $z=\frac{1}{2}L_3$). For the same mode, a hole at $y=\frac{1}{4}L_2$, $z=\frac{1}{4}L_3$ will leave the frequency unchanged.

At least the first of these results can be understood qualitatively from the pattern of

[6] It can be easily verified that the correct final result is obtained by this method even if $q_m \sim \sin \omega t$ or $\cos \omega t$. It is only necessary to go through the derivations from the beginning with a different time dependence.

electric field lines. If the hole is in a plane parallel to the undisturbed electric field (XY or XZ plane), then in the hole the boundary condition $E_z=0$ is relaxed. This is equivalent to a slight extension of the cavity, and therefore the characteristic frequency will be lowered. The electric field near the hole will behave about as indicated in Fig. 3 (exaggerated).

The case of a hole in the YZ plane is not quite so clear. A possible interpretation is that the hole reduces the value of E_z to one-half and that therefore the cavity behaves as if it were smaller, giving a higher frequency. However, a closer investigation of the fields near the hole seems necessary to understand the effects properly.

It will be noted that γ is purely real. This means that only the frequency is changed but not the energy dissipation. This is in agreement with the result in Section 5c that the intensity of the emitted radiation is of order a^6 rather than a^3. The radiation would therefore contribute a dissipative term of order γ^2 only. However, it must be remembered that our theory does not properly take into account the finite conductivity of the cavity walls. The dissipation in the walls will be increased by the presence of the hole (cf. end of Section 7), the increase being of order $a^2\sim\gamma^{\frac{1}{2}}$ [Eq. (62b)].

10. COUPLED CAVITIES

Let us assume two cavities with one common wall with a hole in it (Fig. 1). The fields in the two cavities will be denoted by subscripts α and β, respectively. E.g., $q_{\alpha m}$ will be the amplitude of the mth normal mode of cavity α. Cavity α shall be on the "left" of the boundary $x=0$, while β extends to positive x. We shall assume that mode 1 of α, and mode 1 of β have nearly equal frequencies, and that the cavities are excited with a frequency

$$\omega\approx\omega_{\alpha 1}\approx\omega_{\beta 1}. \tag{76}$$

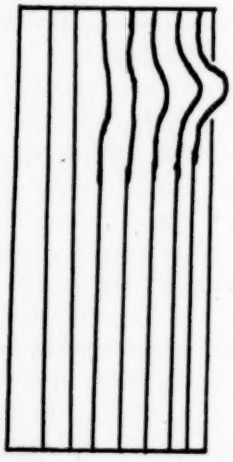

FIG. 3. Electric field lines in a cavity with a small hole in one of the sides parallel to the field. The density of the lines is not proportional to the field strength, the bulge near the hole is exaggerated.

Considering the excitation of cavity β, we have to insert for H_0 and E_{0z} the difference between the fields in cavities α and β. We need only take into account the first mode of each cavity so that, similar to (71):

$$E_{0z}=k_{\alpha 1}q_{\alpha 1}A_{\alpha 1z}-k_{\beta 1}q_{\beta 1}A_{\beta 1z}, \tag{77}$$

$$H_0=-ik_{\alpha 1}q_{\alpha 1}F_{\alpha 1}+ik_{\beta 1}q_{\beta 1}F_{\beta 1}. \tag{77a}$$

We may drop the subscript 1 and refer in the following only to the first mode. Moreover, because of (76), we may put $k_{\alpha 1}=k_{\beta 1}$ in E_{0z} and H_0. Then we obtain the differential equations

$$\begin{aligned}\frac{d^2q_\alpha}{dt^2}+\frac{\omega_0}{Q_\alpha}\frac{dq_\alpha}{dt}+\omega_\alpha^2q_\alpha&=\frac{4}{3}\omega_0^2\frac{a^3}{V_\alpha}(c_{\alpha\alpha}q_\alpha-c_{\alpha\beta}q_\beta),\\ \frac{d^2q_\beta}{dt^2}+\frac{\omega_0}{Q_\beta}\frac{dq_\beta}{dt}+\omega_\beta^2q_\beta&=\frac{4}{3}\omega_0^2\frac{a^3}{V_\beta}(c_{\beta\beta}q_\beta-c_{\alpha\beta}q_\alpha),\end{aligned} \tag{78}$$

with

$$\begin{aligned}c_{\alpha\alpha}=2F_\alpha^2-A_{\alpha z}^2,\quad c_{\beta\beta}=2F_\beta^2-A_{\beta z}^2,\\ c_{\alpha\beta}=c_{\beta\alpha}=2F_\alpha\cdot F_\beta-A_{\alpha z}A_{\beta z}.\end{aligned} \tag{78a}$$

The values of F_α, etc., have to be taken in the hole. In all small terms in (78), ω_α and ω_β have been replaced by ω_0. Equation (78) has exactly the same form as for two conventional coupled circuits and can be treated in the same way.

Let us consider the case of two nearly equal cavities, and let us choose the eigenfunctions so that the normalized magnetic fields F_α and F_β are identical at the common boundary. Then, according to (65c) we have also $A_{\alpha z}=A_{\beta z}$ and therefore

$$c_{\alpha\alpha}=c_{\beta\beta}=c_{\alpha\beta}. \tag{79}$$

We put

$$\gamma=\frac{4}{3}\frac{a^3}{V}(2F_\alpha^2(0)-A_{\alpha z}^2(0)); \tag{79a}$$

then γ is a dimensionless quantity which measures the coupling between the two cavities. Then (78) becomes

$$\frac{d^2q_\alpha}{dt^2}+\frac{\omega_0}{Q}\frac{dq_\alpha}{dt}+\omega_\alpha^2 q_\alpha=\omega_0^2\gamma(q_\alpha-q_\beta),$$
$$\frac{d^2q_\beta}{dt^2}+\frac{\omega_0}{Q}\frac{dq_\beta}{dt}+\omega_\beta^2 q_\beta=\omega_0^2\gamma(q_\beta-q_\alpha). \tag{80}$$

In the last term on the left hand we have kept ω_α and ω_β to allow for a possible lack of tuning. In the other (correction) terms we have put $\omega_\alpha=\omega_\beta=\omega_0$. Neglecting the damping, we find the frequencies of the coupled cavities are given by

$$\omega^2=\tfrac{1}{2}(\omega_\alpha^2+\omega_\beta^2)-\omega_0^2\gamma \pm(\tfrac{1}{4}(\omega_\alpha^2-\omega_\beta^2)^2+\omega_0^4\gamma^2)^{\frac{1}{2}}. \tag{81}$$

If the two cavities are considerably out of tune, i.e., if

$$|\omega_\alpha-\omega_\beta|\gg\omega\gamma, \tag{81a}$$

they will oscillate almost independently, i.e., there will be one proper frequency close to ω_α at which mostly the first cavity is excited while the second has very little excitation, another proper frequency close to ω_β for which the reverse is the case (cf. Section 11).

The cavities are well tuned if

$$|\omega_\alpha-\omega_\beta|\ll\omega\gamma. \tag{81b}$$

The tuning requirement is thus determined by the size of the hole, cf. (79). If $\omega_\alpha=\omega_\beta=\omega_0$, (81) gives

$$\omega^2=\omega_0^2(1-\gamma\pm\gamma). \tag{82}$$

Thus we get one mode whose frequency is the same as for the uncoupled cavities,

$$\omega_1=\omega_0, \tag{82a}$$

and another mode of frequency

$$\omega_2=\omega_0(1-\gamma) \tag{82b}$$

when we neglect terms of order γ^2.

In the first mode we have [cf. (80)]

$$q_\alpha=q_\beta. \tag{83}$$

This means that the tangential magnetic field at the common boundary is continuous, $H_\alpha=H_\beta$, and the same holds for the normal component of the electric field. The hole, then, does not perturb the field at all and the two cavities oscillate as if there were no coupling. This explains immediately why the frequency is not disturbed. Nevertheless, the coupling is important because it determines the phase relations between the two cavities. For the simple case of two rectangular cavities oscillating in their lowest mode we have plotted the electric and magnetic field in Fig. 4. The electric field is assumed to be parallel to the common boundary; it is zero inside the hole just as on the conducting wall.

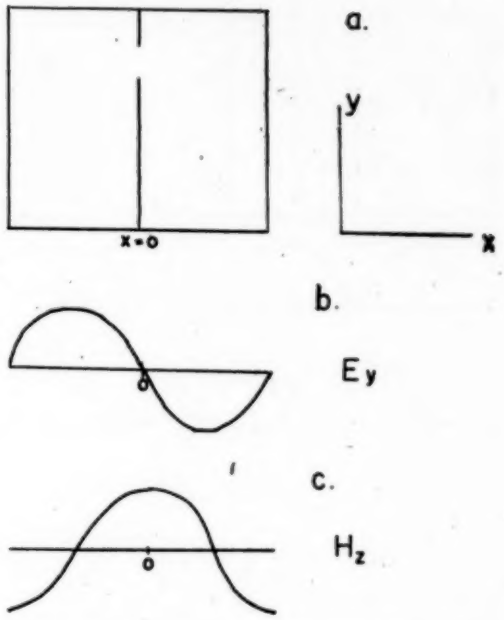

FIG. 4. Field distribution in the first mode ω_1 of two coupled tuned cavities. (a) Cross section of cavities. (b) Electric field E_y, as a function of x (same curve for any value of y). (c) Magnetic field H_z.

The second mode has

$$q_\alpha=-q_\beta. \tag{83a}$$

Then, the magnetic field changes its sign when the boundary between the cavities is crossed (cf. Fig. 5). In the hole, the tangential component of the magnetic field is zero. Around the hole, there is a large perturbation of the field. The tangential electric field has a maximum in the hole. In Fig. 5 we have plotted the electric and magnetic field for the same cavity as Fig. 4, but this time for the second mode, $\omega_2=\omega_0(1-\gamma)$. Since in our case the unperturbed electric field has no component normal to the common boundary, γ is positive, and the frequency is reduced by the coupling.

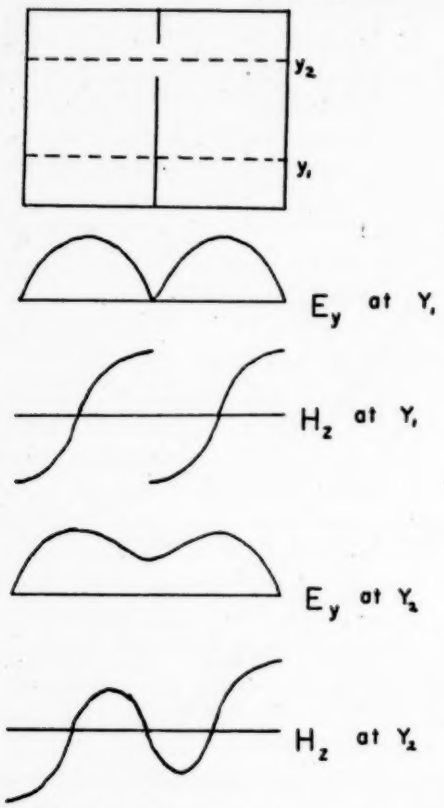

FIG. 5. Field distribution in the second mode ω_2 of the cavities of Fig. 4. (a) Cross section [identical with 4(a)]. (b) Electric field E_y along a line not going through the hole ($y=y_1$). (c) Magnetic field along the same line. (d) Electric field E_y along a line $y=y_2$ going through the hole. (e) Magnetic field along the same line.

This corresponds to the fact that now, so to speak, the two cavities oscillate as a whole; the electric field E_y no longer goes to zero in the hole but has a secondary maximum there, thus beginning to develop towards the field corresponding to the lowest mode in the larger cavity $\alpha+\beta$ with the central boundary absent (Fig. 6). This mode is entirely different from that of the separate cavities (Fig. 4); therefore in mode 2 (which is the lowest harmonic for the larger cavity of Fig. 6), a large perturbation field exists around the hole which varies rapidly in space. The energy dissipation is likely to be considerably greater than for mode 1 (cf. Section 7, end).

The coupling coefficient γ depends sensitively on the position of the hole, and may be zero for certain positions. For example, in our case of two rectangular cavities with an electric field in the y direction, the only field component which is not zero on the common boundary ($x=0$) is H_z. If we do not restrict our attention to the lowest mode, we have for the normalized field:

$$F_z=\frac{m_1/L_1}{[(m_1/L_1)^2+(m_3/L_3)^2]^{\frac{1}{2}}} \times 2\cos\frac{\pi m_1 x}{L_1}\sin\frac{\pi m_3 z}{L_3}, \quad (84)$$

where m_1 and m_3 are positive integers not zero and $L_1L_2L_3$ are the lengths of the three edges of a single cavity. Then we have for the coupling coefficient [cf. (79)]

$$\gamma=\frac{8}{3}\frac{a^3}{L_1L_2L_3} \times\frac{(m_1/L_1)^2}{(m_1/L_1)^2+(m_3/L_3)^2}4\sin^2\frac{\pi m_3 z}{L_3}. \quad (84a)$$

The last factor possesses nodes on the common boundary, $x=0$, for any value of m_3 greater than one. In general, large coupling is desired; then the hole should be put at or near the maximum of γ.

If definite phase relations between the oscillations of the two cavities are required, the correct frequency must be selected. It seems possible to predict, in many cases at least, whether the desired phase relation corresponds to the higher or the lower frequency. In any case, two coupled

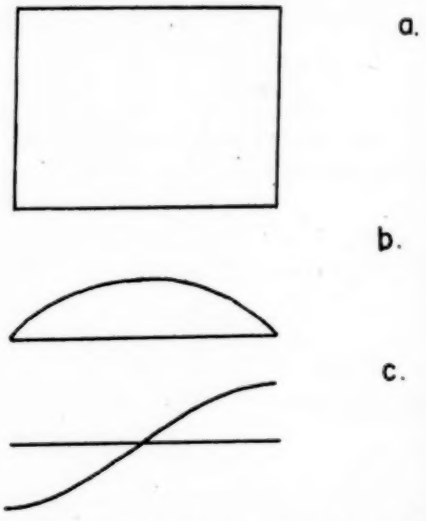

FIG. 6. (a) The cavity of Figs. 4 and 5 without the wall in the middle. (b) Electric field distribution E_y in the lowest mode of the cavity a. (c) Magnetic field in the same mode.

cavities which are exactly tuned will give two modes with a relative frequency difference of the order $\gamma \sim a^3/V$, the phase relations between the fields in the two cavities being opposite for the two modes.

11. STEP-UP COUPLING

It is often desirable to couple two circuits in such a way that the field in β is stronger than in α. This can be achieved for two exactly tuned cavities if they are coupled asymmetrically, *viz.*, by a hole which is in a position corresponding to a high value of the normalized field F_α and to a low one of F_β (Fig. 7). Then the excitation may be expected to adjust itself so that the magnetic field is continuous in the hole, *viz.*:

$$H_\alpha = q_\alpha F_\alpha \text{ (hole)} = H_\beta = q_\beta F_\beta \text{ (hole)}, \quad (85)$$

where F_α, F_β are the normalized fields. Then, since F_β is made small compared with F_α, we get $q_\beta \gg q_\alpha$ as desired.

We shall show in this section that (85) is actually fulfilled for *one* of the two modes, *viz.*, that having the same frequency as the uncoupled cavities. There is, however, another mode in which the field in β is smaller than in α. Furthermore, we are going to show that two slightly mistuned cavities coupled symmetrically will give the same results as two exactly tuned cavities coupled asymmetrically.

The problem of cavities coupled asymmetrically in the way indicated in Fig. 7 can be easily treated by Eqs. (78) to (78c). It is convenient to define the "displaced frequencies."

$$\begin{aligned} \omega_\alpha'^2 &= \omega_\alpha^2 - (4/3)\omega_0^2 c_{\alpha\alpha}(a^3/V), \\ \omega_\beta'^2 &= \omega_\beta^2 - (4/3)\omega_0^2 c_{\beta\beta}(a^3/V), \end{aligned} \quad (86)$$

and the coupling coefficient

$$\gamma = (4/3)\omega_0^2 c_{\alpha\beta}(a^3/V), \quad (86a)$$

the volumes of the two cavities having been assumed equal. Then, if ω denotes the actual frequency, (78) becomes

$$\begin{aligned} (\omega_\alpha'^2 - \omega^2)q_\alpha + \omega_0^2\gamma q_\beta &= 0, \\ \omega_0^2\gamma q_\alpha + (\omega_\beta'^2 - \omega^2)q_\beta &= 0, \end{aligned} \quad (87)$$

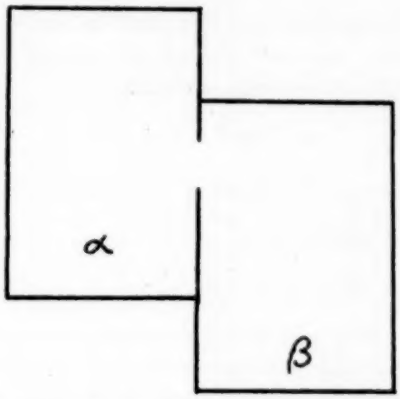

FIG. 7. Asymmetrical coupling of the two tuned cavities.

which has the two solutions

$$\omega_{1,2}^2 = \tfrac{1}{2}(\omega_\alpha'^2 + \omega_\beta'^2) \pm [\tfrac{1}{4}(\omega_\alpha'^2 - \omega_\beta'^2)^2 + \omega_0^4\gamma^2]^{\frac{1}{2}}. \quad (87a)$$

The ratio of the excitation amplitudes q_α, q_β is:

$$\frac{q_\beta}{q_\alpha} = \lambda \pm (1+\lambda^2)^{\frac{1}{2}}, \quad (88)$$

the upper sign corresponding to the higher frequency ω_1, the lower sign to ω_2. λ is given by

$$\lambda = \frac{\omega_\beta'^2 - \omega_\alpha'^2}{2\gamma\omega_0^2} \approx \frac{\omega_\beta' - \omega_\alpha'}{\gamma\omega_0}. \quad (88a)$$

We shall now make the simplifying assumptions [cf. (78a), (78b)] that A_x is zero in the hole, and that F_α and F_β have the same direction. These assumptions are fulfilled by the cavities treated in Section 10 where $A = A_y$ and F is in the z direction everywhere on the common boundary of the cavities. Then we may write [cf. (78a), (78b)]

$$c_{\alpha\alpha} = c_{\alpha\beta}\mu, \quad (89)$$

$$c_{\beta\beta} = c_{\alpha\beta}/\mu, \quad (89a)$$

with

$$\mu = F_\alpha/F_\beta. \quad (89b)$$

Then (86) reduced to

$$\begin{aligned} \omega_\alpha'^2 &= \omega_\alpha^2 - \gamma\omega_0^2\mu, \\ \omega_\beta'^2 &= \omega_\beta^2 - \gamma\omega_0^2/\mu, \end{aligned} \quad (90)$$

and (88a) becomes

$$\lambda = (\mu^2 - 1)/2\mu. \quad (90a)$$

If we now assume that the cavities were originally tuned, i.e., $\omega_\alpha=\omega_\beta=\omega_0$, we obtain, by inserting (90), (90a) into (87a) and (88)

$$\omega_1^2=\omega_0^2, \tag{91}$$

$$\omega_2^2=\omega_0^2[1-\gamma(\mu^2+1)/\mu], \tag{91a}$$

$$(q_\beta/q_\alpha)_1=\mu, \tag{92}$$

$$(q_\beta/q_\alpha)_2=-1/\mu. \tag{92a}$$

We find thus one mode ω_1 of frequency equal to the unperturbed frequencies, and another mode ω_2 of lower frequency. In the first mode the ratio of the amplitudes of excitation q_β/q_α is such that the magnetic fields in the hole are matched, in agreement with Eq. (85). This mode, then, gives the desired stepping up of the amplitude from α to β. However, in the second mode, the ratio of the q's is reversed (in addition, the phase relation is opposite). Caution is therefore necessary to excite the coupled cavities with the correct frequency in order to obtain the desired results. (For frequencies other than ω_1 and ω_2, beats will occur.)

Moreover, we see that it is not at all necessary to couple the cavities unsymmetrically in order to step up the excitation from α to β. It is only important [cf. (88)] that λ be different from zero, i.e. [cf. (88a)], *that* ω_α' and ω_β' *be different*. This can be achieved either by making $\omega_\alpha=\omega_\beta$ but $c_{\alpha\alpha}\neq c_{\beta\beta}$; then we have two tuned but unsymmetrically coupled cavities as treated in (89) to (92a). But it can equally well be achieved by two slightly mistuned and symmetrically (or also unsymmetrically) coupled cavities; then $\omega_\alpha\neq\omega_\beta$ and $c_{\alpha\alpha}=c_{\beta\beta}$ so that

$$\lambda=\frac{\omega_\beta-\omega_\alpha}{\gamma\omega_0}. \tag{93}$$

In either case, the more strongly excited cavity is the one whose ω' is closer to the actual frequency.

Slow Mesons in Cosmic Radiation

S. V. CHANDRASHEKHAR AIYA AND R. C. SAXENA
Cosmic-Ray Research Unit, Indian Institute of Science, Bangalore, India
(Received June 2, 1944)

Using Bhabha's method, the intensity of slow mesons is determined. The intensity of mesons of energy greater than 10^8 ev and less than 2.7×10^8 ev is 6 percent of the total meson intensity. Near the geomagnetic equator, the total intensity increases by 54.4 percent, the meson intensity by 21.7 percent, and the electron intensity by 162.1 percent when the altitude increases from 3100 ft. to 7200 ft. The electron intensity, which is about 25 percent of the meson intensity at 3100 ft., increases to 50 percent of the meson intensity at 7200 ft.

THE behavior of cosmic-ray electrons is explained satisfactorily by the cascade theory. On the basis of this theory, very few, if any, of the primary electrons incident on the top of the atmosphere can reach sea level. Experimental results, however, indicate clearly that a significant part of cosmic radiation at sea level, and even at great depths, consists of electrons. It is believed that most of these electrons in the lower atmosphere arise out of meson decay.[1] The test of the correctness of any theory of meson decay ultimately depends on an accurate determination of the proportions of electrons and mesons in cosmic radiation.

One method of determining this ratio of electron to meson component is by a statistical examination of the tracks obtained in a Wilson chamber. For high accuracy, a large number of photographs are required. It is not possible to distinguish mesons of all energies from electrons by their tracks in a Wilson chamber.[2,3] Stray photographs of low energy mesons may be obtained but this fact can only be utilized to indicate the presence of very low energy mesons in cosmic radiation.

The second method is by the use of Geiger counters. With a counter telescope, the total electron and meson intensity can be determined. (The intensity of protons, etc., if present can be considered to be negligible and, for purposes of this paper, will be included in the meson component.) By introducing an absorber like lead between the counters of a counter telescope, electrons can be absorbed and the meson intensity determined. The difference between the total and the meson intensity gives the electron intensity. This method has its obvious handicaps. If small thicknesses of lead are used to absorb electrons, some of the high energy electrons are not cut out and are included in the meson component. If large thicknesses of lead are used for the purpose, low energy mesons are absorbed and they will be included in the electron component. Mesons of energy, 3×10^8 ev can, according to Rossi and Greisen,[4] penetrate 21.76 cm of lead. Lead absorbers of thickness up to this value have been used in estimating the meson intensity. Therefore, it is usual to call all mesons of energy less than 3×10^8 ev as "slow" mesons. Mesons of higher energy are called "fast" mesons.

If the absorption curve of slow mesons is determined, it is possible to obtain the ratio of electron to meson intensity from the results of the usual experiments. Besides, the intensity of slow mesons of different energies can be determined. In recent years, attempts have been made to determine the intensity of slow mesons. Auger[5] and Greisen[6] make use of the differential absorption of electrons and mesons in the same equivalent thickness of different materials like lead, carbon, aluminum, etc., to arrive at an estimate of slow meson intensity. Auger's results are subject to considerable error. Greisen's analysis is a very elegant one, and his results are in good agreement with the results reported in this paper. But such a method is an indirect one, and an

[1] Euler and Heisenberg, Ergeb. d. exact. Naturwiss **17**, 1 (1938).
[2] Blackett, Proc. Roy. Soc. **A165**, 11 (1938).
[3] S. H. Anderson and C. D. Nedermeyer, Phys. Rev. **51**, 884 (1937); **54**, 88 (1938).
[4] B. Rossi and K. Greisen, Rev. Mod. Phys. **13**, 240 (1941).
[5] P. Auger, Phys. Rev. **61**, 684 (1942).
[6] K. Greisen, Phys. Rev. **63**, 323 (1943).

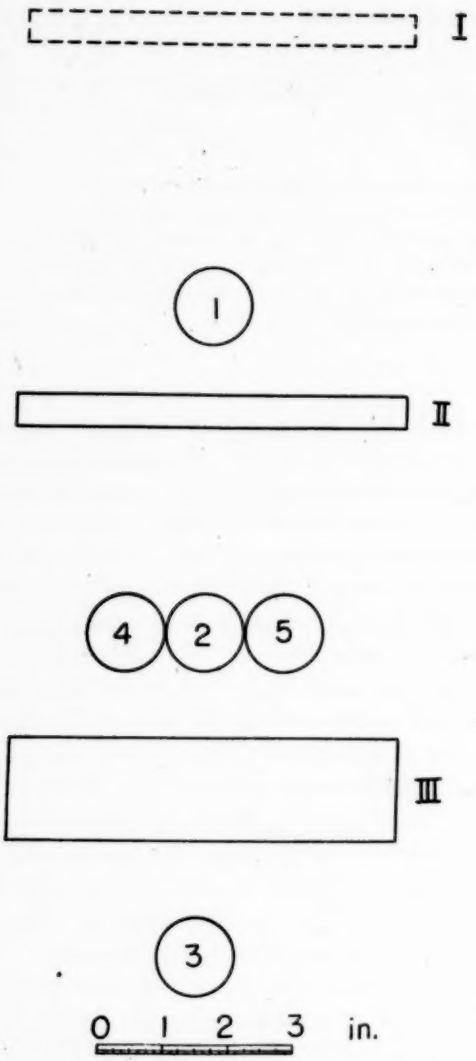

FIG. 1. Arrangement of counters.

estimate of the slow meson intensity by a more direct method is, therefore, desirable. Schein and others[7] and Sarabhai[8] cut out electrons while measuring the meson intensity by not recording particles which produce showers in small thicknesses of lead. It has been pointed out by Bhabha[9] that a good fraction of low energy electrons produce showers of very few particles, often only of one and that this fraction cannot be cut out by just not registering the shower-producing particles. This fact has also been experimentally verified by one of us.[10]

Any attempt to estimate the intensity of "slow" mesons must be free from the objections indicated. In addition, the results must be free from the effects of side showers. In order to minimize the possibility of chance coincidences, it is desirable to avoid a double coincidence system. Impelled mainly by these considerations, an attempt has been made here to give a report of a new approach to the problem made possible by the method developed by Bhabha.[9]

EXPERIMENTAL ARRANGEMENTS

A counter telescope designed on the basis of Bhabha's method is shown in Fig. 1. Counters *1*, *2*, and *3* were connected in coincidence and form the vertical telescope. Counters *4* and *5* were connected in parallel and in anticoincidence with counters *1*, *2*, and *3*. The pulse was taken from the copper cylinders of counters *4* and *5* for anticoincidence and these counters were shielded. Each of the counters used was 6 in. long and $1\frac{1}{4}$ in. in diameter. This type of counter telescope gives results free from the effect of side showers.[10]

Positions I, II, and III were for placing lead. High energy electrons produce showers in 1.25 cm of lead placed in position II and are cut out by the shower particles tripping the counters 4 or 5 or both. Lead in position II loses a good deal of its significance when lead is placed in position I. Even under such conditions, lead was placed in position II, for purposes of maintaining identical geometrical conditions for all measurements. Lead placed in position III was for absorbing low energy electrons. Lead completely covering the solid angle was placed in position I for absorption measurements. As penetrating non-ionizing cosmic rays form a negligible fraction of the total

TABLE I. Measurements of intensity.

No.	Lead I	Lead II	Lead III	Total lead	Counts	Time in hours	Counts/ hour
1	—	—	—	—	4497	38	118±1.2
2	—	—	5.25	5.25	5058	52	97.3±0.9
3	—	1.25	4.00	5.25	7680	81	94.8±0.7
4	5.2	1.25	4.00	10.45	4826	52	92.8±0.9
5	9.3	1.25	4.00	14.55	6905	76	90.8±0.7
6	13.7	1.25	4.00	18.95	5430	61	89.0±0.8

[7] Schein, Jesse, and Wollan, Phys. Rev. **58**, 1027 (1940).
[8] V. Sarabhai, Proc. Ind. Acad. Sci. **19**, 37 (1944).
[9] H. J. Bhabha, Proc. Ind. Acad. Sci. **19**, 23 (1944).
[10] Chandrashekhar Aiya, Proc. Ind. Acad. Sci. (in print).

intensity[11] and as scattering can be considered negligible for purposes of the accuracy aimed at, this choice of position I for lead in absorption measurements is not objectionable.

The measurements taken are given in Table I. From observations 3, 4, 5, and 6, a graph is drawn and is shown in Fig. 2. This gives the absorption curve of slow mesons. It is extrapolated to zero thickness by dashes.

DISCUSSION OF THE RESULTS

No. 1 represents the total intensity. If we assume that No. 3 gives the meson intensity, the electron intensity is 23.5 ± 1.9 counts/hour. Therefore, at Bangalore (geomagnetic latitude $=3.3°$N and altitude $=3100$ ft. above mean sea level) cosmic rays consist of 80 percent mesons and 20 percent electrons. It will be clear from the discussion which follows that No. 3 gives the meson intensity to an accuracy of 2 percent and, therefore, the ratio of electron to meson intensity stated above is accurate to within 2 percent. The difference between No. 2 and No. 3 represents the fraction of electrons cut out by splitting the lead as suggested by Bhabha. It is 10.6 ± 6.8 percent of the total electron intensity.

According to Bhabha, the fraction of electrons which penetrates 5.25 cm of lead and which does not produce adequate showers to trip counters 4 or 5 or both is at the most 2 percent of the total electron intensity. Therefore 0.47 count in 94.8 ± 0.7 counts/hour represents the electron intensity registered as mesons in observation No. 3. This is within the limits of accuracy of the measurement. Consequently it is correct to say that No. 3 gives the intensity of mesons which penetrate 5.25 cm of lead. Observations 4, 5, and 6 give the meson intensities to a still better degree of accuracy. Therefore, the graph drawn with these observations represents the absorption curve of slow mesons. By making use of this graph and the curves of Rossi and Greisen,[4] we can deduce the intensity of mesons within different energy limits. Thus, it is found that slow mesons of energy greater than 10^8 ev and less than 2.7×10^8 ev is about 6 percent of the total meson intensity.

[11] Janossy and Rochester, Proc. Roy. Soc. **A181**, 399 (1943).

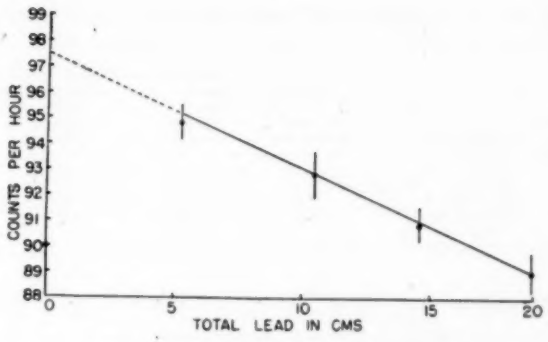

FIG. 2. Absorption of slow mesons.

One of us[12] has reported something abnormal in the behavior of mesons of energy in the region of 3×10^8 ev, and the subject is yet under investigation. It is, therefore, not proposed to go into the question of mesons of this energy here.

The curve given by Euler and Heisenberg[1] for the energy spectrum of mesons shows that mesons of energy less than 10^8 ev constitute about 2 percent of the total meson intensity. The smallness of this fraction is due to the collision loss of mesons which increases rapidly below this energy. If our graph were extrapolated to zero thickness, which is clearly not justified for the reason just mentioned, we would find that about 2.7 percent of the total meson intensity consists of mesons of energy less than 10^8 ev. This is natural. The meson absorption curve for thicknesses of lead less than 5.25 cm must fall off towards the X axis, i.e., must be below the dotted line in Fig. 2. Although it is not conclusive, it may be pointed out that the point for 5.25 cm of lead having a higher statistical accuracy than others is below the line and can be taken to indicate that the curve does fall off as is to be expected. The percentage of mesons absorbed in the first 5.25 cm of lead, i.e., of energy less than 0.93×10^8 ev is very small (less than 2 percent of the total intensity), and it is difficult to estimate this small quantity by obtaining readings of meson intensity only within the first 5.25 cm of lead as any method used does not entirely cut out electrons and the fraction of electrons not cut out increases with decreasing thickness of lead.

It would, however, be possible to obtain an *upper limit* to the meson intensity at 1.25 cm of lead by omitting the 4.0 cm of lead absorber.

[12] Chandrashekhar Aiya, Nature **153**, 375 (1944).

TABLE II. Comparison of results at 3100 ft. and 7200 ft.

Component	Intensity at 3100 ft. in counts/hr. (Bangalore)	Intensity at 7200 ft. in counts/hr. (Ootacamund)	Percent increase corresponding to a rise in altitude of 4100 ft.
Total	118.3	182.7	54.4
Mesons	94.8	121.1	21.7
Electrons	23.5	61.6	162.1

Sarabhai[8] has done this with 2 cm of lead. Since the intensity of the soft component is 25 percent of the meson intensity at Bangalore and we can gauge from the figures given by Bhabha[9] that very roughly about 10 percent of the soft component passes through this thickness without producing showers, it follows that the soft component would contribute about $2\frac{1}{2}$ percent to the meson reading so obtained. But it is quite clear that such a measurement would have no meaning if taken at an altitude of 7200 ft. where the soft component intensity is 50 percent of that of the hard component and, at still greater altitudes the error would be much greater.

The graph in Fig. 2 holds for the solid angle of the counter telescope used by us and indicated in Fig. 1, which is drawn to scale. For different solid angles, a part of the mesons at least will have traveled through different distances in the atmosphere and suffered different degrees of losses of energy. It is, therefore, not quite impossible that the absorption curve may be *slightly* different for different solid angles, but this is a matter for investigation.

The counter telescope used for experiments at Ootacamund (geomagnetic latitude = 1.7°N and altitude = 7200 ft. above mean sea level) by one of us (S.V.C.) was quite similar to the one used here. The results at the two places can be utilized to indicate the variation of the different components with altitude near the geomagnetic equator. This is done in Table II.

The electron intensity which is about 25 percent of the meson intensity at 3100 ft. increases to 50 percent of the meson intensity at 7200 ft.

ACKNOWLEDGMENT

The authors desire to thank Professor H. J. Bhabha, F.R.S., for his interest and encouragement and the trustees of the Sir Dorabji Tata Trust who have financed this research. The authors' thanks are also due to Dr. K. R. Ramanathan of the India Meteorological Department for the loan of Professor Millikan's counters which were tested and used in the experiments reported here.

The Paths of Ions and Electrons in Non-Uniform Magnetic Fields*

NORMAN D. COGGESHALL AND MORRIS MUSKAT
Gulf Research and Development Company, Pittsburgh, Pennsylvania
(Received July 29, 1944)

The integration of the Lorentz force equations to give electron or ion paths has been reduced to simple quadratures for systems in which the electric field is zero and the magnetic field is a function of one Cartesian or cylindrical coordinate. For several interesting types of magnetic field variation the quadrature can be carried through analytically; and even for complicated magnetic fields, or such as are known only empirically, the numerical integration can be effected without difficulty. From general considerations of the functions involved, it is possible to determine the extension and periodicity of the orbits for any set of initial conditions. The representation used is also convenient for obtaining information regarding the dispersion and focusing characteristics of the trajectories, some of which have unique properties of promise for use in specific instrument design such as mass spectrometers, beta-ray spectrographs, etc. Schematic designs of such instruments are proposed, and a discussion is given of their advantages and special properties.

INTRODUCTION

THE focusing and dispersing properties of a uniform magnetic field have been applied with great success in recent years to mass spectrographs, mass spectrometers, and β-ray spectrometers, as well as to more special uses.

Many instruments have been built and described which employ either 180° focusing or the focusing properties of sector-shaped magnetic fields. All are generically of the same class, of which the 360° analyzer is the only one having the property of perfect focusing. However, it is not practical to attempt to utilize the latter because of the physical impossibility of locating both a source of ions and a collector at exactly the same place.

In contrast to the wide application of the focusing and dispersing properties of a uniform magnetic field, little use has been made of the focusing properties of non-uniform fields. The problem of calculating paths, dispersion, and focusing characteristics for a general type of non-uniform field is one of great mathematical difficulty.[1] There are, however, some cases of considerable physical interest in which the magnetic field varies in a fairly simple manner, and for which it is possible to calculate exactly the paths, dispersion, and focusing properties.

In all cases where the pole faces of a magnet or an electromagnet are symmetrically located with respect to a median plane, the magnetic field at any point on the plane will be perpendicular to it. For an ion or electron moving in this plane the problem of describing its path reduces to a two-dimensional one.

For those cases where the magnetic field in the median plane can be expressed as a function only of one Cartesian coordinate, or as a function only of the radial distance from the axis of symmetry, for cases of circular symmetry, the equations describing the ion or electron paths admit of a simple treatment. Even if the dependence of the field upon such coordinates is known only experimentally, the complete solution may be determined by simple numerical integration to an accuracy equal to that with which the magnetic field is known. The general theory will be outlined and discussed below and certain cases analyzed in detail.

GENERAL THEORY

The Lorentz force equation for a charged particle in the presence of a magnetic field H and the absence of an electric field, is:

$$\mathbf{F}=\frac{e}{c}\mathbf{v}\times\mathbf{H}, \tag{1}$$

where e = charge of particle in e.s.u.; c = velocity of light; and $\mathbf{v}$ = velocity of particle in cm/sec.

* For an abstract of a preliminary report of this work see Phys. Rev. **65**, 352 (1944).

[1] While there is a vast literature on the general optics of charged particles moving in magnetic fields, applications of the type considered here do not appear to have been given specific treatment.

In addition to Eq. (1), we have the equation:

$$(\dot{x})^2+(\dot{y})^2+(\dot{z})^2=v^2=\text{const.},$$

since the magnetic field can do no net work on the charge. For the case of both Cartesian and cylindrical coordinates, we may, without loss of generality, consider H to be parallel to the z direction, so we then have:

$$\begin{aligned}(\dot{x})^2+(\dot{y})^2&=v^2,\\ (\dot{r})^2+(r\dot{\theta})^2&=v^2.\end{aligned} \tag{2}$$

Here v refers to the resultant velocity in a plane perpendicular to the z direction.

Written out in terms of Cartesian components, Eq. (1) becomes:

$$\begin{aligned}\ddot{x}&=-\frac{e}{mc}H\dot{y},\\ \ddot{y}&=\frac{e}{mc}H\dot{x},\end{aligned} \tag{3}$$

where H is the magnitude of the magnetic field, or:

$$\begin{aligned}\ddot{r}-r(\dot{\theta})^2&=-\frac{e}{mc}Hr\dot{\theta},\\ \frac{1}{r}\frac{d}{dt}(r^2\dot{\theta})&=\frac{e}{mc}H\dot{r},\end{aligned} \tag{4}$$

in cylindrical coordinates. Equations (3) and (4) cannot in general be solved explicitly if H is a function of both coordinates, as the equations will not be linear.

If, however, H is a function of only one coordinate, the equations are solvable in a simple manner, as follows:

Referring to Eq. (3), we may assume, without loss of generality, for this case that $H=H(x)$, and define a function $f=f(x)$ as:

$$f=\frac{e}{vmc}\int H(x)dx. \tag{5}$$

Integrating the second of Eqs. (3), we get:

$$\dot{y}=\frac{e}{mc}\int Hdx=vf. \tag{6}$$

However, from Eq. (2), we have:

$$\dot{x}=\pm(v^2-\dot{y}^2)^{\frac{1}{2}}=\pm v(1-f^2)^{\frac{1}{2}}. \tag{7}$$

Since $dy/dx=\dot{y}/\dot{x}$, we may combine Eqs. (6) and (7) to obtain:

$$\frac{dy}{dx}=\pm\frac{f}{(1-f^2)^{\frac{1}{2}}}, \tag{8}$$

or:

$$y=\pm\int\frac{fdx}{(1-f^2)^{\frac{1}{2}}}. \tag{9}$$

We see thus that when H is a function of only one Cartesian coordinate, the equations of the trajectory are reduced to the simple differential equation (8), which permits immediate integration. When the magnetic field is a function only of the cylindrical coordinate r, the treatment is similar. For this case we define a function $g(r)$ by:

$$g(r)=\frac{e}{vmc}\frac{1}{r}\int rH(r)dr. \tag{10}$$

Integrating the second of Eqs. (4), we find

$$r\dot{\theta}=\frac{e}{rmc}\int rHdr=vg. \tag{11}$$

Using Eqs. (2) and (11), we have:

$$\dot{r}=\pm v(1-g^2)^{\frac{1}{2}}, \tag{12}$$

and since $r\dot{\theta}/\dot{r}=r(d\theta/dr)$, we have from Eqs. (11) and (12):

$$\frac{d\theta}{dr}=\pm\frac{1}{r}\frac{g}{(1-g^2)^{\frac{1}{2}}}, \tag{13}$$

or:

$$\theta=\pm\int\frac{gdr}{r(1-g^2)^{\frac{1}{2}}}. \tag{14}$$

We see, therefore, that the treatments for the Cartesian case and for the cylindrical case are similar, and involve essentially nothing more than two quadratures, either or both of which may be done numerically if not analytically.

Equations (5) and (10), which define f and g, will each contain a constant of integration. These constants may be used to specify the slope of the trajectory for any desired value of x or r by direct application of Eqs. (8) or (13). The

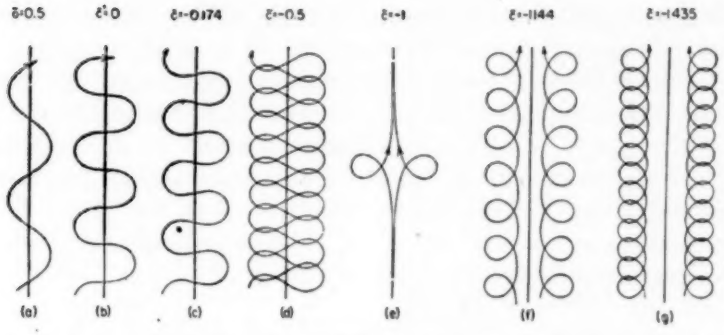

FIG. 1. Various types of orbits possible in a linearly varying magnetic field, determined by the parameter c, defined by Eq. (17).

other initial condition, namely, the starting position in the x, y or θ, r plane is taken care of by the constant of integration in Eqs. (9) or (14).

The fact that the equations defining the trajectories involve the radical $(1-f^2)^{\frac{1}{2}}$ for the Cartesian case, and the radical $(1-g^2)^{\frac{1}{2}}$ for the cylindrical case, has a very simple and significant interpretation. Since these radicals must represent real magnitudes for any physically possible trajectory, we see immediately that the extension of the trajectory in the x or r coordinate will be given by $-1\leqslant f\leqslant 1$, or $-1\leqslant g\leqslant 1$, respectively.

If the equation

$$1-f^2=0 \tag{15}$$

has two real roots, the implication is that orbits will be bounded as regards extension in the x coordinate. This follows from the fact that dy/dx will be infinite for x equal to a real root of Eq. (15). If Eq. (15) has only one real root the orbit is one that comes in from and returns to plus or minus infinite values of x. For the case where Eq. (15) has two real roots the orbit will be either completely closed, or have an infinite extension in the y direction. The latter orbits will be of periodic behavior and will, in general, consist of an infinite number of loops of cycloidal appearance.

The same general considerations as just discussed for the case of Cartesian coordinates also hold for the cylindrical system and figures demonstrating them will follow in the detailed discussion of specific cases.

In the calculations for the paths in either coordinate system the quantity e/vmc enters as a fundamental constant of the motion. It is convenient to define this quantity as a, i.e.:

$$a=e/vmc.$$

In the discussion of many cases of physical interest it is more expedient to refer to the potential of a charged particle as a measure of its energy rather than to its velocity. We may, therefore, write a as

$$a=\frac{1}{c}\left(\frac{150e}{mV}\right)^{\frac{1}{2}},$$

where V is the potential of the particle in volts or its energy in electron volts.

LINEARLY VARYING MAGNETIC FIELD

The linear magnetic field may be defined by:

$$H=bx, \tag{16}$$

the origin being placed where the field is zero. From Eq. (5) it follows that:

$$f=a\int Hdx=\frac{abx^2}{2}+c, \tag{17}$$

where c is the constant integration and $a=e/vmc$, as previously indicated.

Introducing the change of variables:

$$\bar{y}=\left(\frac{ab}{2}\right)^{\frac{1}{2}}y;\quad \bar{x}=\left(\frac{ab}{2}\right)^{\frac{1}{2}}x, \tag{18}$$

so that $\bar{y}$ and $\bar{x}$ are dimensionless, we find Eq. (8) takes the form:

$$\frac{d\bar{y}}{d\bar{x}}=\pm\frac{c+\bar{x}^2}{[(1-c-\bar{x}^2)(1+c+\bar{x}^2)]^{\frac{1}{2}}}. \tag{19}$$

By inspection of the right side of this equation it is clear that if it is to represent a real trajectory $\bar{c}<1$. Moreover, the nature of the real trajectories will be determined by the magnitude of $\bar{c}$, as follows:

For $1>\bar{c}>-1$, the trajectories will all cross the y axis, with a slope $=\bar{c}/[1-\bar{c}^2]^{\frac{1}{2}}$. They will be symmetrical with respect to y, and the ion or electron will oscillate periodically between fixed values of x. If $1>\bar{c}>0$, the trajectory will simulate a sinusoidal oscillation with a monotonic increase or decrease in y, as illustrated in Fig. 1(a). When $\bar{c}=0$, the trajectories will have an inflection when crossing the y axis, as shown in Fig. 1(b), so that the change in y is still monotonic as the path is traversed. For $0>\bar{c}>-1$, the variation in y is no longer monotonic, and the trajectory tends to form loops as it oscillates between the fixed limits in x [cf. Fig. 1(c)]. At $\bar{c}=-0.46$, the loops of consecutive periods on either side of the x axis touch, and become interlaced [cf. Fig. 1(d)] as $\bar{c}$ is still further decreased. Complete overlapping of the loops, so as to form a single figure-eight trajectory, develops for $\bar{c}=-0.6522$. For the range $-0.6522>\bar{c}>-1$, they are again similar to a network of interlaced figure of eight oscillations, which spread out and ultimately degenerate into the single pair of split loops when $\bar{c}=-1$, as shown in Fig. 1(e).

In this whole range of $\bar{c}$ the maximum extension of the oscillations, which may be obtained from the roots of Eq. (15), continually increases as $\bar{c}$ decreases, according to the equation:

$$\bar{x}_{\max}=(1-\bar{c})^{\frac{1}{2}}. \tag{20}$$

Moreover, from the roots of Eq. (15) the width of the oscillations is given by:

$$\bar{x}_{\max}-\bar{x}_{\min}=(1-\bar{c})^{\frac{1}{2}}-(1+\bar{c})^{\frac{1}{2}}. \tag{21}$$

The detailed calculation of the trajectories may be readily made by introducing the substitution: $\bar{x}=(1-\bar{c})^{\frac{1}{2}}\cos\Phi$ and reducing the integral of Eq. (19) to standard elliptic integral form. It is thus found that:

$$\bar{y}=\frac{1}{\sqrt{2}}[2E(k)-K(k)-2E(\Phi,k)+F(\Phi,k)], \tag{22}$$

where $E(\Phi,k)$, $F(\Phi,k)$ are the incomplete elliptic integrals corresponding to E, K. It is to be noted that Eq. (21) applies directly only to the quarter-cycle of a single complete oscillation passing through the origin. Its continuation to both preceding and following elements of the trajectory is easily made by simple shifting of the initial point and appropriate changes in sign, as dictated by Eq. (19).

The wave-length, along y, of the oscillations is given by:

$$\lambda=2\sqrt{2}\,|2E(k)-K(k)|, \tag{23}$$

where K and E are the complete elliptic integrals of the first and second kind, the modulus $k=[(1-\bar{c})/2]^{\frac{1}{2}}$, and the absolute value is used to take care of the change in sign of the enclosed term at $\bar{c}=-0.6522$. λ is expressed in the same units as $\bar{y}$. It will be noted that $\lambda=4\bar{y}(\bar{x}_{\max})$.

In the limiting case $\bar{c}=-1$, integration of Eq. (19) readily gives:

$$\pm\bar{y}=\frac{1}{\sqrt{2}}\log\frac{\sqrt{2}+(2-\bar{x}^2)^{\frac{1}{2}}}{\bar{x}}-(2-\bar{x}^2)^{\frac{1}{2}},\quad \bar{x}>0, \tag{24}$$

from which were plotted the trajectories of Fig. 1(e).

The trajectories of greatest practical interest are those for $\bar{c}<-1$. The general features of these can also be deduced by direct inspection of Eq. (19). These no longer cross the $\bar{y}$ axis, but split up into two identically shaped individual trajectories, for each value of $\bar{c}$, placed symmetrically on the two sides of the $\bar{y}$ axis. Each of these is composed of a series of interlaced or separated loops extending from

$$|\bar{x}_{\max}|=(-1-\bar{c})^{\frac{1}{2}},\quad \text{to}\quad |\bar{x}_{\min}|=(1-\bar{c})^{\frac{1}{2}}, \tag{25}$$

and hence recede from the $\bar{y}$ axis as $\bar{c}$ is decreased. The overlapping of the loops increases as $\bar{c}$ decreases [cf. Figs. 1(f) and 1(g)]. The wave-length is given by:

$$\lambda=2(1-\bar{c})^{\frac{1}{2}}\left[\left(1-\frac{k^2}{2}\right)K-E\right] \tag{26}$$

where now the modulus $k=[2/(1-\bar{c})]^{\frac{1}{2}}$. A plot of λ and $(\bar{x}_{\max}-\bar{x}_{\min})$ *vs.* $\bar{x}_{\max}$ is given in Fig. 2. It will be seen that both λ and the lateral range of the oscillations decrease as $\bar{x}_{\max}$ increases—i.e., as the trajectories are confined to regions of greater magnetic field. This is, of course, to be expected from elementary considerations.

For calculating the exact shape of the trajectories for this range of $\bar{c}$, the integral of Eq. (19) may again be transformed into standard elliptic integrals, by setting $\bar{x}^2=\cos 2\Phi-\bar{c}$ with the result that:

$$\bar{y}=(1-\bar{c})^{\frac{1}{2}}[E(\Phi,k)-(1-\tfrac{1}{2}k^2)F(\Phi,k)]. \quad (27)$$

As remarked above, orbits of the type illustrated in Figs. 1(f) and 1(g) are of greatest interest. This is because of the practical difficulty of obtaining a field variation in which H is an odd linear function of x. Also, for any restricted region of an actual non-uniform field H may be considered linear with distance to a first approximation. The orbits calculated from Eq. (19) will therefore represent the actual paths of the ions or electrons to at least a first approximation. Application of such orbits as in Figs. 1(g) and 1(f) to magnetic focusing instruments will be discussed below.

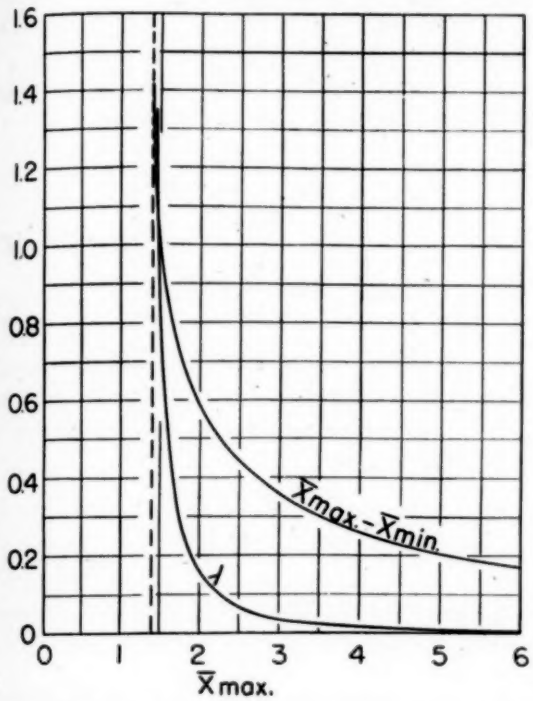

FIG. 2. The variation of the dimensionless wave-length λ and lateral range $\bar{x}_{max}-\bar{x}_{min}$, of periodic orbits in a linearly varying magnetic field, with the maximal extension $\bar{x}_{max}$.

EXPONENTIALLY VARYING FIELD

For the case of an exponentially varying field in Cartesian coordinates we may write H as:

$$H=H_0e^{bx},$$

where H_0 is the magnitude of the field strength at the value of x chosen as the origin of the coordinate system, and b is the constant determining the rate of change of H with x.

The function f then becomes:

$$f=\frac{aH_0}{b}e^{bx}+\bar{c}, \quad (28)$$

where $\bar{c}$ is a constant of integration which may be adjusted to give the orbit the desired slope at any value of x. For this case Eq. (8) becomes:

$$\frac{dy}{dx}=\pm\frac{\frac{aH_0}{b}e^{bx}+\bar{c}}{\left\{1-\left(\frac{aH_0}{b}e^{bx}+\bar{c}\right)^2\right\}^{\frac{1}{2}}}. \quad (29)$$

By making the transformation:

$$v=\frac{aH_0}{b}e^{bx},$$

Eq. (29) may be readily integrated to give the general solution:

$$y=\pm\frac{1}{b}\left\{\sin^{-1}(v+\bar{c})-\frac{\bar{c}}{(1-\bar{c}^2)^{\frac{1}{2}}}\right.$$
$$\left.\times\log\left[\frac{\{1-(v+\bar{c})^2\}^{\frac{1}{2}}+(1-\bar{c}^2)^{\frac{1}{2}}}{v}-\frac{\bar{c}}{(1-\bar{c}^2)^{\frac{1}{2}}}\right]\right\}, \quad (30)$$

when $|\bar{c}|<1$, and

$$y=\pm\frac{1}{b}\left\{\sin^{-1}(v+\bar{c})\right.$$
$$\left.+\frac{\bar{c}}{(\bar{c}^2-1)^{\frac{1}{2}}}\sin^{-1}\left[\frac{1-\bar{c}^2-\bar{c}v}{v}\right]\right\}, \quad (31)$$

when $\bar{c}<-1$.

For the special case, $\bar{c}=-1$,

$$y=\pm\frac{1}{b}\left\{2\sin^{-1}\left(\frac{v}{2}\right)^{\frac{1}{2}}+\frac{(2v-v^2)^{\frac{1}{2}}}{v}\right\}.$$

Orbits defined by Eq. (30) are illustrated by (a), (b), and (c) in Fig. 3. In each case the particle enters the magnetic field from $x=-\infty$, or, physically speaking, from a region of very weak H. As must obviously be the case, each orbit is symmetrical about the x axis. Each path

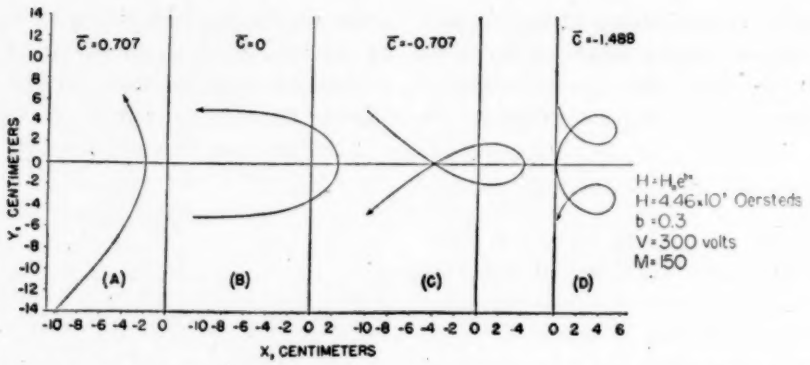

FIG. 3. Various types of orbits possible in an exponential magnetic field, determined by the parameter $\bar{c}$, defined by Eq. (28). All orbits are for singly-charged particles of 300-ev energy and 150 atomic mass units. *Note:* The second line of the explanation at the right of the cut should read $H_0=4.46\times10^3$ oersteds.

is shown in the neighborhood of $H=H_0$, which establishes the origin of the x coordinate. Orbits (a), (b), and (c) represent a particle of 150 atomic mass units and 300 electron volts energy coming from $x=-\infty$ with initial slope of 1, 0, and -1, respectively. Orbit (d) is for a particle of the same mass and energy but so projected in the x, y plane that the path has infinite slope at $x=0$.

Orbits of the type illustrated by (b) are of particular interest, as the total extension in the y direction depends only on the exponential coefficient b, and is independent of the values of the mass and velocity of the ion or electron. This fact is immediately seen upon setting $\bar{c}=0$ in Eq. (30), which physically corresponds to the particles approaching regions of stronger magnetic field with an original direction parallel to the x direction. This situation affords an opportunity for perfect focusing, further discussion of which will be given later.

If the ion or electron is introduced into the magnetic field in such a manner that its subsequent orbit is bounded in the x direction, it will be of the type illustrated by (d), Fig. 3. This type of orbit is also of interest as regards application to focusing problems, and will be discussed in more detail below. For this type trajectory the total width in the x dimension is readily calculated from the separation of the roots of Eq. (15).

The behavior of orbits of type (d) as regards their total width in the x direction and their wave-length λ can be discussed in a manner similar to that used for the linearly varying field. For a particular value of $\bar{c}$ we have:

$$x_{\max}=\frac{1}{b}\log\frac{b}{aH_0}(1-\bar{c}),$$
$$x_{\min}=\frac{1}{b}\log\frac{b}{aH_0}(-1-\bar{c}). \tag{32}$$

As $\bar{c}<-1$ for all such orbits the arguments of the logarithms will be positive. We may express $x_{\max}-x_{\min}$ as a function of $x_{\max}$ by eliminating $\bar{c}$ in Eqs. (32) to give:

$$x_{\max}-x_{\min}=x_{\max}-\frac{1}{b}\log\left(e^{bx_{\max}}-\frac{2b}{aH_0}\right),$$

which behaves qualitatively like the similar quantity for the linear case which is plotted in Fig. 2.

From the nature of Eq. (31) it is seen that the orbits are periodic with a wave-length of λ given by:

$$\lambda=\frac{2\pi}{b}\left(1+\frac{\bar{c}}{(\bar{c}^2-1)^{\frac{1}{2}}}\right),$$

or

$$\lambda=\frac{2\pi}{b}\left[1+\frac{1-\frac{aH_0}{b}e^{bx_{\max}}}{\left[\left(1-\frac{aH_0}{b}e^{bx_{\max}}\right)^2-1\right]^{\frac{1}{2}}}\right]. \tag{33}$$

RADIAL FIELDS

Although Eqs. (10) and (14) make it possible to determine the path of a charged particle in

any field of circular symmetry, the functional variation of H we shall discuss here is given by:

$$H=\frac{H_0}{r^{n+1}},$$

where $n>-1$, and H_0 is a constant. It will, of course, be physically impossible for H to obey this relation for values of r including the origin. However, it will suffice if the relation is obeyed within certain limits of r, or for values of r above a definite radius.

For this case (and $n\neq 1$) $g(r)$ becomes:

$$g=\frac{aH_0}{(1-n)r^n}+\frac{\bar{c}}{r}, \tag{34}$$

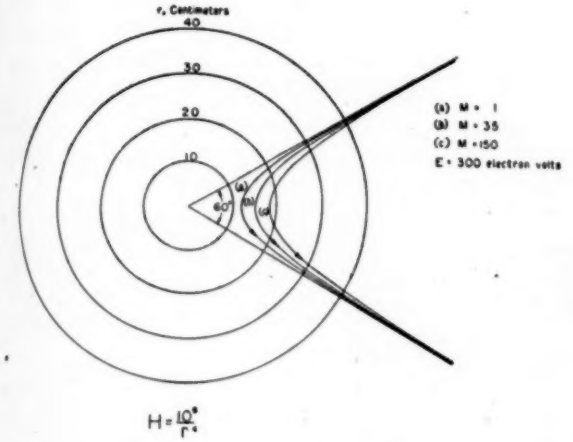

FIG. 4. Non-periodic orbits with common asymptotes in a radially varying magnetic field, with $n=3$, for singly-charged particles of 300 ev and various atomic masses M.

and we have:

$$\theta=\pm\int\left(\frac{aH_0}{(1-n)r^{n+1}}+\frac{\bar{c}}{r^2}\right)\Big/\left[\left(1-\left\{\frac{aH_0}{(1-n)r^n}+\frac{\bar{c}}{r}\right\}^2\right)\right]^{\frac{1}{2}}dr. \tag{35}$$

Equation (35) is in general more amenable to numerical than analytical treatment.

One case of important physical interest does, however, admit of simple analytical treatment. This is that of charged particles approaching regions of strong H from regions of weak H, and with an initial direction along a radius vector and toward the axis of symmetry.

Before discussing the solution for this particular case it will be of interest to examine the general features of the orbits defined by Eq. (35). It will be noted first, from Eq. (34), that if $n>0$ $(\neq 1)$, $g(r)\to 0$ as $r\to\infty$. Hence by Eq. (13), $r(d\theta/dr)\to 0$ as $r\to\infty$. This means that for $n>0$ $(\neq 1)$ all orbits entering the field from infinity must be asymptotically parallel to a radius vector. As all radii vectors are equivalent, these orbits may thus be considered as necessarily entering the field parallel to the polar axis $(\theta=0)$. If, in particular, these orbits asymptotically degenerate into linear paths at a normal distance h from the polar axis, i.e., so that

$$r\sin\theta=h,$$

then it follows that in the limit as $r\to\infty$,

$$r\frac{d\theta}{dr}\longrightarrow-\frac{h}{r\cos\theta}\sim-\frac{h}{r}\sim g.$$

By Eq. (34), such an asymptotic behavior can be satisfied only if $n>1$ and $\bar{c}=-h$. Accordingly, in inverse power radial fields for which $n>1$, $-\bar{c}$ is simply the normal distance from the polar axis with which the particle enters the field. The restriction $n>1$ evidently means that if $n<1$ there can be no orbits which enter the field along linear paths parallel to a radius vector. If $0<n<1$ the value of $\bar{c}$ can be determined by fixing $r(d\theta/dr)$ at a given radius, computing g from Eq. (13), and then $\bar{c}$ by Eq. (34).

If $n=0$, the orbits entering or leaving the field at $r=\infty$ will asymptotically cut the radii vectors at an angle $\sin^{-1}aH_0$. For this case the integral for $\theta(r)$ can be readily evaluated to give expressions similar to Eqs. (30) and (31). If $-1<n<0$ the orbits will be restricted to finite regions of the (r,θ) plane, and be bounded by a maximum value of r.

Returning to the special case $n>1$, and also assuming that the particle actually enters the field along the polar axis $(\bar{c}=0)$, we may readily integrate Eq. (35) and obtain:

$$\theta=\pm\frac{1}{n}\sin^{-1}\frac{aH_0}{(1-n)r^n}. \tag{36}$$

Orbits of this type for $n=3$ are plotted in Fig. 4. It is interesting to note that these orbits

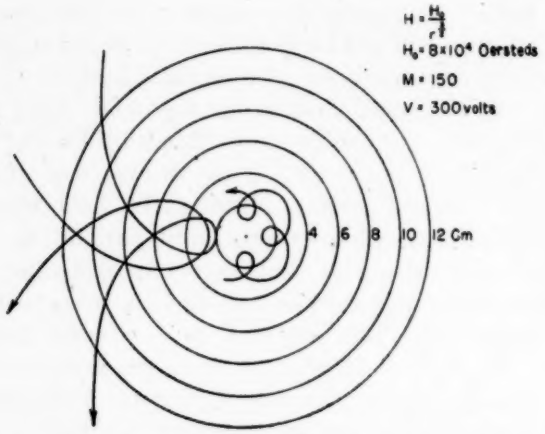

FIG. 5. Typical orbits in a radially varying magnetic field, with $n=\frac{1}{2}$, which are convex to the origin at $r=r_{\min}$.

have an angular spread independent of mass and energy, quite analogous to those for the exponentially varying field in Cartesian coordinates, shown in Fig. 3(b). These trajectories will be discussed in more detail later.

Other types of orbits obtainable in radial fields are illustrated in Figs. 5 and 6. In Fig. 5 the particles are so projected, in a field with $n=\frac{1}{2}$, that the paths are convex with respect to the origin at the points of closest approach. These were determined by numerical integration of Eq. (35), after choosing the point of minimum approach and determining the corresponding value of $\bar{c}$. In a similar manner were computed the orbits shown in Fig. 6, which are concave to the origin at the points of closest approach. In Fig. 5 may be seen a periodic or cyclic type of orbit of the same general properties as possessed by the periodic orbits in the Cartesian coordinate cases.

For periodic orbits, the equation

$$1-g^2=0,$$

which is analogous to Eq. (15) for the Cartesian case, must have two real and unequal roots. For a given value of n and $\bar{c}$ it is possible to determine whether the resulting orbit will be periodic or not by determining whether g will pass through both the values -1 and $+1$ as r is increased.

APPLICATIONS

In addition to illustrating the application of the general method of calculation given above to specific cases, the orbits previously discussed have certain characteristics of promise for instruments such as mass spectrometers, mass spectrographs, beta-ray spectrometers, etc. For example, Fig. 7 illustrates the orbits of ions of masses 1, 35, and 150, expressed in atomic mass units, and all having a kinetic energy of 300 electron volts. The type of equation which describes these orbits, as well as that of Fig. 3(b), is obtained by setting $\bar{c}=0$ in Eq. (30), giving

$$y=\pm\frac{1}{b}\sin^{-1}\left(\frac{aH_0}{b}e^{bx}\right). \tag{37}$$

From these figures and Eq. (37) it is seen that all such ions which initially approach an exponentially varying field in the direction of increasing H will turn and recede in an opposite direction. Also, the ions will recede asymptotically to a line displaced from their original line of approach by an amount independent of the momentum of the particle and depending only upon the exponential coefficient b.

With this arrangement all ions beginning at a point where the field is negligibly weak, and with a direction toward regions of increasing H, will return to a single point where the magnetic field is also negligibly weak.

A schematic diagram of an instrument which might be designed to utilize this type of orbit may be seen in Fig. 8. The design shown here is for a mass spectrometer, but with obvious changes it could be used as a beta-ray spec-

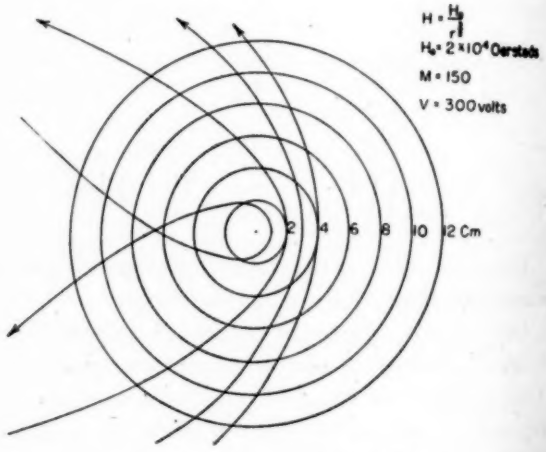

FIG. 6. Typical orbits in a radially varying magnetic field, with $n=\frac{1}{2}$, which are concave to the origin at $r=r_{\min}$.

trometer, or for special studies such as electron or ion scattering, etc. By adjusting an aperture or slit, located as shown in Fig. 8, the instrument may be made to focus either only particles of a narrow momentum range, such as ions of a definite mass, or particles of a wide momentum or energy range. An idea of the amount of dispersion taking place at the region of the slit may be obtained by inspection of Fig. 7, which represents paths of physical interest. The separation in the turning points of any two ion beams of momentum parameters a_1 and a_2 may be readily calculated from Eq. (37), which gives

$$x_1-x_2=\frac{1}{b}\log\frac{a_2}{a_1}=\frac{1}{b}\log\frac{(mv)_1}{(mv)_2}.$$

For an instrument such as this, it would be necessary for the magnetic field to vary in an exponential manner only over a region large enough to encompass the dimensions of the instrument.

In designing pole pieces of an electromagnet so as to obtain an exponentially varying field, the first requirement is to have the dimension in the direction perpendicular to the x, y plane long compared to the size of the instrument desired, so as to allow the field to be accurately two dimensional. Next it would be necessary that the design be such that the three equations:

$$\frac{\partial^2 M}{\partial x^2}+\frac{\partial^2 M}{\partial z^2}=0, \tag{38}$$

$$\left.\begin{aligned}\frac{\partial M}{\partial x}&=0\\ \frac{\partial M}{\partial z}&=H_0e^{bz}\end{aligned}\right\}\text{at } z=0, \tag{39}$$

be satisfied, where M represents magnetostatic potential.

A solution of Eq. (38), which also satisfies Eqs. (39), is:

$$M=\frac{H_0e^{bz}}{b}\sin bz.$$

If the magnetostatic potentials of the opposing pole faces are $+M_0$ and $-M_0$, respectively, we have the equation:

$$z=\pm\frac{1}{b}\sin^{-1}\left(\frac{bM_0e^{-bz}}{H_0}\right), \tag{40}$$

to define the contours of the pole faces.

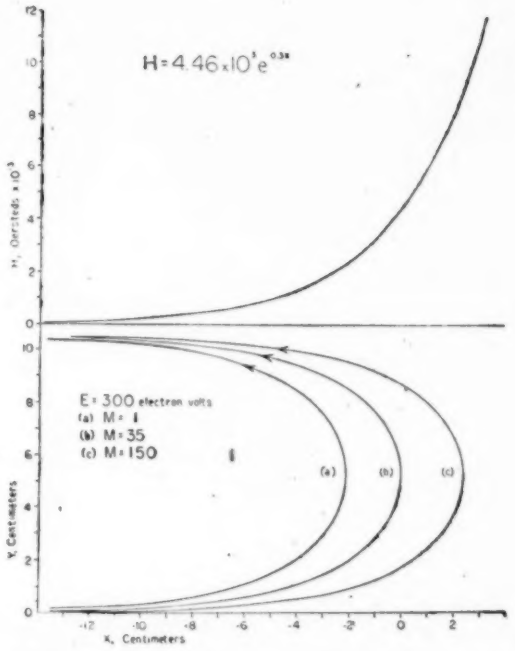

FIG. 7. Typical orbits with common asymptotes in an exponentially varying magnetic field.

In Fig. 4 are seen orbits which, for the case of radially symmetric fields, are analogous to the orbits just discussed. They are defined by Eq. (36). It is obvious that since these orbits are analogous to those of Figs. 7 and 5, the same focusing properties could be applied in corresponding instruments. As for the Cartesian exponential field, the optics of the radial field system is such that ions originating at a point where the field is negligibly weak, i.e., at "infinity," and with an initial velocity toward the center of symmetry of the field, will be all focused, regardless of their individual momenta, at another single point. This point will be on a radius vector which is rotated from the line of approach radius vector by an angle which depends only upon the exponent in the equation defining the variation of H with r. It must also be remembered that orbits with this property are possible only for $n>1$ in Eq. (34).

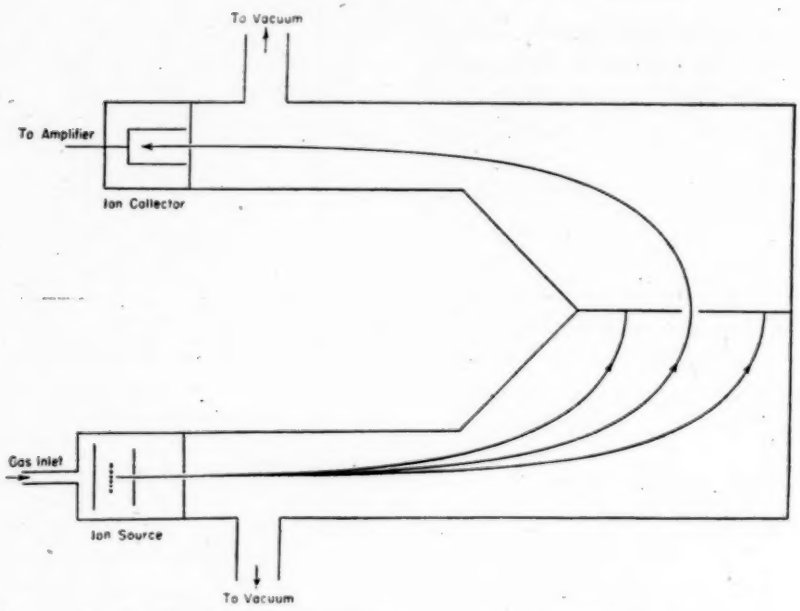

FIG. 8. Schematic diagram of a magnetic analyzer utilizing an exponentially varying magnetic field.

A schematic diagram of an instrument designed to employ this type of orbit is shown in Fig. 9. An idea of the dispersion obtainable may be seen from Fig. 4, and may be expressed analytically as:

$$\frac{\Delta r}{r}=\frac{1}{n}\frac{\Delta a}{a}=\frac{1}{n}\frac{\Delta(mv)}{mv},$$

where r and $r+\Delta r$ are the turning radii for particles of parameters a and $a+\Delta a$. Fields which decrease with radial distance with the same functional form have been successfully used in the betatron.[2]

To obtain the field variation just discussed the circularly symmetric pole faces must be so shaped that the following equations are satisfied:

$$\frac{\partial}{\partial r}\left(r^2\frac{\partial M}{\partial r}\right)+\frac{1}{\sin\Phi}\frac{\partial}{\partial\Phi}\left(\sin\Phi\frac{\partial M}{\partial\Phi}\right)=0, \quad (41)$$

$$\left.\begin{aligned}\frac{\partial M}{\partial r}&=0,\\ \frac{1}{r}\left(\frac{\partial M}{\partial\Phi}\right)&=\frac{H_0}{r^{n+1}},\end{aligned}\right\}\Phi=\pi/2, \quad (42)$$

[2] D. W. Kerst, Phys. Rev. **60**, 47 (1941).

where M is the magnetostatic potential, and Φ is the latitude angle.

Setting $\mu=\cos\Phi$, and $M=(H_0/r^n)F(\mu)$, we find $F(\mu)$ must satisfy the Legendre equation:

$$(\mu^2-1)\frac{d^2F}{d\mu^2}+2\mu\frac{dF}{d\mu}-n(n-1)F=0.$$

To fix the specific solutions for $F(\mu)$ applying to the present problem, the equivalents of Eq. (42), namely: $F(0)=0$; $F'(0)=-1$, are to be imposed on the functions $F(\mu)$.

If we let the opposite pole faces have magnetostatic potential $+M_0$ and $-M_0$, the contours will be given by:

$$r=\left\{\frac{H_0F(\theta)}{\pm M_0}\right\}^{1/n}, \quad (43)$$

the sign of $F(\theta)$ automatically changing in passing from one pole face to the other, so as to ensure a real value for r.

As the region over which it is desired that H have the specified functional form will not extend to the origin, the pole faces in this region need not follow the shape required by Eq. (43). This physical observation also obviates the apparent difficulty arising from the singularities at $\mu=\pm1$ in the functions $F(\mu)$ when n is not an integer.

The periodic or cyclic orbits of the type in Figs. 1(f) and (g), and Fig. 3(d), also have properties of promise for focusing applications. One of the most important of these properties is illustrated in Fig. 2. This shows how the wave-length λ, or spatial periodicity of the orbit, decreases with increasing x_{max}. This curve was drawn for H varying linearly with x, but its character will be essentially the same for other functional variations wherein H increases monotonically with increasing x.

An application of such orbits may be made to the design of a mass spectrometer or similar instrument, as indicated schematically in Fig. 10. In such an instrument the ions are so accelerated by a suitable source that they are projected into the magnetic field with the x_{max} for each orbit lying within the exit slits.

In Fig. 11 may be seen a schematic diagram of the state of affairs at the exit slits S_1 and S_2 of such an ion source. We shall assume that the potential distribution in the source is such that all ions acquire essentially the same potential, and that the last two slits, i.e., S_1 and S_2 in Fig. 11, are at the same potential and are operative only in defining the emergent beam. Moreover, these slits will be assumed to be geometrically similar and located identically as regards the x coordinate.

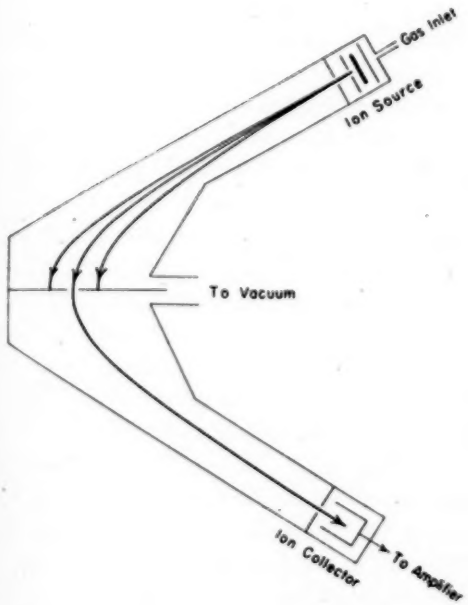

FIG. 9. Schematic diagram of a magnetic analyzer utilizing a radially varying magnetic field.

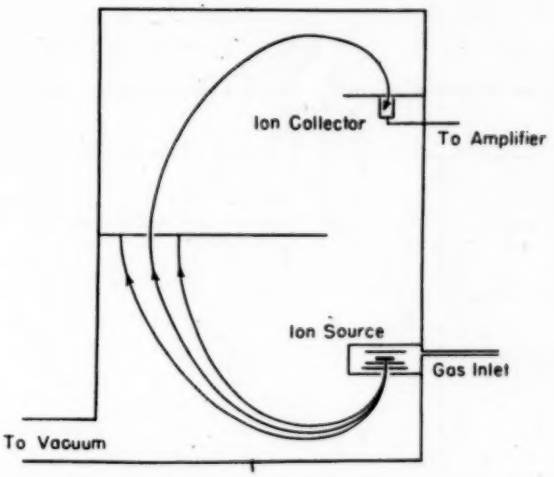

FIG. 10. Schematic diagram of a magnetic analyzer utilizing periodic orbits in an exponentially varying magnetic field.

With such an arrangement all ions leaving the exit slits will follow orbits having approximately the same radius of curvature at the slits, differing only in the angle of emergence.

In Fig. 11, (c) represents the orbit with the smallest value of x_{max} that can emerge from the slits, (d) that with the largest value of x_{max}, while (a) and (b) represent the paths of greatest divergence in angle of emergence. As S_1 and S_2 cover the same range along the x direction, the x_{max} will be the same for orbits (a) and (b).

Consider now the orbits (a), (b), (c), and (d) when the particles have advanced by exactly one wave-length in their periodic paths. Since (a) and (b) have the same x_{max}, they will have the same λ. Therefore a slit S_1' of the same width as S_1 and located at the same values of x, but displaced along y by a distance λ, will collect both (a) and (b). That is, (a) and (b) will pass through S_1' with exactly the same geometry as they passed through S_1. As orbit (c) has a smaller value of x_{max}, it will be displaced slightly upward, relative to (a) and (b). Conversely, as orbit (d) has a larger value of x_{max}, its λ will be smaller than that of (a) and (b). Hence as it passes through S_1' it will be displaced slightly downward relative to (a) and (b).

This shifting of the orbits (c) and (d) relative to (a) and (b) means that the latter will limit the lateral spread in the ion beam when entering S_1' just as they defined the angular spread in

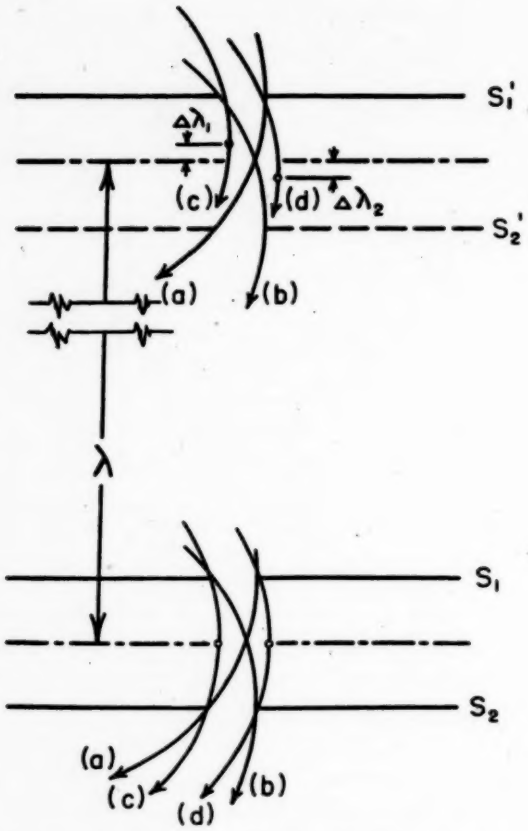

FIG. 11. Schematic diagram of the geometry of the orbits at the source and collector slits of an instrument of the type shown in Fig. 10.

the beam emerging from the source slits S_1 and S_2. In the design of such an instrument, as schematically shown in Fig. 10, the final two exit slits will correspond to S_1 and S_2, and the collector slit should be of the same dimensions as S_1 and located at the same distance along the x axis. The separation along the y axis between S_1 and S_1' will then define the wave-length λ for which the above-described focusing process applies. Any ion may be made to follow an orbit of any chosen λ by properly fixing the accelerating voltage in the ion source. It is to be noted that for the collector the one slit S_1' is sufficient. S_2' is shown in Fig. 11 only to indicate the geometry of the orbits at this point. If the instrument is to be used so that only ions of one mass or of small range of momentum are to pass the collector, an aperture or slit may be introduced, as shown in Fig. 10.

The calculation of the dispersion at the region of the selector slit or aperture is easily made by application of Eq. (29). For this purpose it is convenient to choose the origin at the exit slits of the ion source. Then, since $dy/dx=\infty$ at the ion source ($x=0$), $\bar{c}$ must have the value

$$\bar{c}=1-\frac{aH_0}{b}.$$

At the turning point of the orbit, i.e., at the plane of the selecting slits, $dy/dx=\infty$ again, so that:

$$\bar{c}+\frac{aH_0}{b}e^{bx}=-1,$$

giving for the location of the turning point:

$$x=\frac{1}{b}\log\left\{1-\frac{2b}{aH}\right\}.$$

For two ion beams of momentum parameters a_1 and a_2 the turning points will then be separated by a distance:

$$(x_1-x_2)=\frac{1}{b}\log\left\{\frac{1-\dfrac{2b}{a_1H}}{1-\dfrac{2b}{a_2H}}\right\}.$$

A similar type of instrument could be designed using the periodic orbits in a radial field, such as shown in Fig. 5. Similar considerations with respect to focusing and the use of selector slits will be applicable to such instruments. Finally, it should be noted that the above discussion has referred to particles moving in the median plane between the pole pieces generating the magnetic field. Hence in practical designs account will have to be taken of the perturbations in the orbits of particles off the median plane owing to the curvature of the magnetic flux lines.

The authors are indebted to Mr. N. F. Kerr for assistance in preparing the drawings and making the calculations, and to Dr. P. D. Foote, Executive Vice President of The Gulf Research and Development Company, for permission to publish this paper.

PHYSICAL REVIEW VOLUME 66, NUMBERS 7 AND 8 OCTOBER 1 AND 15, 1944

The X-Ray *K*-Absorption Edge for Sodium and the Secondary Structure* Accompanying the Edge for the Sodium Halides

KENNETH C. RULE†
Institute of Physics, University of Upsala, Upsala, Sweden
(Received May 24, 1944)

A high vacuum, curved-crystal spectrograph was constructed for study of the *K*-absorption edge and the accompanying secondary absorption structures for evaporated films of sodium metal and of the sodium halides. Comparison of the $K\beta_2$ emission line and the *K*-absorption edge for metallic sodium indicates that the $K\beta_2$ line is not semi-optical in origin, but represents a transition from the $M_{II\ III}$ valence shell. The energies of the 1*s*, 2*p*, and 3*p* electron energy levels for metallic sodium are calculated. The energy difference between the $K\beta_1$ emission line and the *K*-absorption edge for Na^+ in NaCl is found to be 5 ev greater than the difference between the $K\beta_5$ emission line and the *K*-absorption edge for Cl^- in NaCl. Strong secondary absorption patterns are found on the short wave-length side of the *K*-absorption edge for Na^+ in the sodium halides. The distances from the various maxima and minima in the pattern to the main edges are correlated with the lattice distances and satisfactory agreement with Kronig's hypothesis is found except for the features near the main edge in the spectrum for NaF. Fair agreement is also found in comparison of these patterns with the patterns accompanying the *K*-absorption edge for K^+ in the potassium halides, as determined by Brewington.

IN Kronig's[1] hypothesis of the origin of the secondary absorption structure which often appears on the short wave-length side of x-ray absorption edges of elements in crystal lattices, it is supposed that the final energies of the electrons ejected in any given direction through the lattice are restricted by the occurrence of Bragg-like reflections in the periodic potential field in the lattice. Owing to the overlapping of the effects of reflections from numerous sets of planes, actual calculations of secondary absorption patterns have not been carried out rigorously even for the simplest lattice types, although approximate calculations have been attempted.[2] A number of simple deductions from the theory have, however, been subjected to extensive experimental tests in which metals and binary alloys have been employed as absorbing media. Most of these experimental data have been found to be harmony with the theory,[1–4] but in the case of alloys the crystal lattice is not always the decisive factor in determining the secondary structure.[5]

The most serious discrepancies with theory have been encountered in the case of absorption edges of elements in ionic lattices.[6–8] Since Kronig's theory applies only to the motion of a "free electron," discrepancies in the structure located within twenty-five or fifty electron volts from the edge can be attributed to effects of the field about the ion from whose x-ray level the electron is ejected. Stephenson[9] has found, however, that the ionic charge and the ionic group from which the electron is ejected affect the secondary structure at least as far as 100 ev from the edge.

It has been deduced from Kronig's theory that for an ion occurring in a succession of ionic lattices of a given type and varying lattice distances, the difference in energy between a given feature in the secondary absorption structure and the main edge should be inversely proportional to the square of the lattice distance. Since the

* This work was carried out at the Institute of Physics of the University of Upsala in Sweden and was aided by the G. Hilmer Lundbeck fellowship of the American-Scandinavian Foundation.
† Author's present address: South Charleston Division, Westvaco Chlorine Products Corporation, South Charleston, West Virginia.
[1] R. de L. Kronig, Zeits. f. Physik **70**, 317 (1931); D. Coster and J. Veldkamp, Zeits. f. Physik **74**, 191 (1932).
[2] R. de L. Kronig, Zeits. f. Physik **75**, 191 (1932); C. Kurylenko, J. de phys. et rad. **1**, 133 (1940).
[3] D. Coster and H. Levy, Physica **6**, 44 (1939).
[4] J. H. Munier, J. A. Bearden, and C. H. Shaw, Phys. Rev. **58**, 537 (1940).
[5] J. A. Bearden and H. Friedman, Phys. Rev. **58**, 387 (1940).
[6] Vola P. Barton and Geo. A. Lindsay, Phys. Rev. **46**, 362 (1934); S. Yoshida, Sci. Papers Tokyo Inst. Phys.-Chem. Research **38**, 272 (1941).
[7] D. Coster and G. H. Klamer, Physica **1**, 145 (1933).
[8] G. P. Brewington, Phys. Rev. **46**, 861 (1934).
[9] S. I. Stephenson, Phys. Rev. **58**, 873 (1941).

absolute separation depends only on the lattice and is independent of the energy of the main edge, it is evident that the relative separation of the secondary absorption structure from the main edge should be greatest for an edge of long wave-length. Previous experiments with ionic lattices as absorbing media have been confined to K and L edges of comparatively short wave-lengths. Early indications were that secondary absorption becomes weaker for elements of low atomic number,[7] but recent work has succeeded in bringing out the secondary structure in the K edges for metallic Al and Mg.[4] In view of the realizable gain in resolving power, it therefore seemed worth while to attempt to study the structure accompanying an edge at still longer wave-length, such as the K edge for Na in its halides. A second advantage in a study of the sodium halides is that the lattice constants increase as much as 1.4-fold from the fluoride to the iodide, a range which should make it possible to draw significant conclusions as to the correctness of Kronig's hypothesis in respect to dependence of the separation from the main edge on lattice distance.

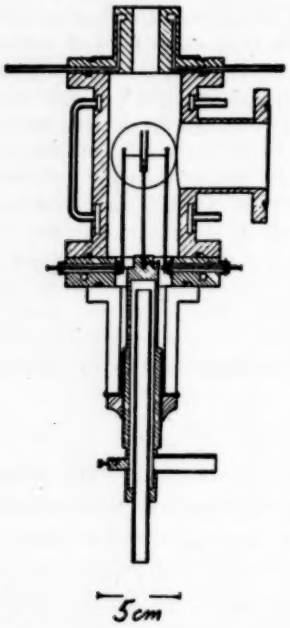

FIG. 1. Cross section of the cathode-ray oven.

Aside from the objective of studying the secondary absorption structure, a determination of the wave-length of the main sodium edge, in which the various halides and also sodium metal were used as absorption films, was of interest in itself. The only previous determination found in the literature was an incidental measurement by Johnson,[10] who employed sodium chloride as the absorption film.

EXPERIMENTAL

Apparatus

The apparatus and technique employed were designed to avoid excessively long times of exposure in bringing out absorption edges and secondary absorption structure in the region of long wave-lengths. The curved-crystal type of spectrograph was therefore chosen. A further advantage of this instrument is that the edge and the absorption spectrum extending far out from the edge can be photographed on a single plate at one setting of the crystal. The high vacuum extended throughout the spectrograph so that no diaphragm was required over the mouth of the x-ray tube. Since thin, homogeneous absorption films would further contribute to lowering the exposure time, the films were formed by condensation on an aluminum foil of vapors produced by heating the materials in a cathode-ray oven. In order that films of metallic sodium as well as of the sodium halides could be deposited and employed without exposure to air, the cathode-ray oven was constructed as an integral part of the high vacuum spectrograph.

The cathode-ray oven, Fig. 1, was essentially a water-cooled, metal, x-ray tube mounted at the bottom of the crystal chamber. The anticathode was designed[11] to minimize loss of heat by radiation and was thimble-shaped in order to facilitate heat transfer to the material to be evaporated. A loop of tungsten wire at ground potential was the source of an electron beam of 1 to 200 milliamperes which served to maintain the crucible at 600 to 1400°C when a potential of 1 to 20 kv was applied to the anticathode. The oven was evacuated through a flexible metal tube connected directly to the oil-diffusion pump.

The x-ray tube was a water-cooled, metal tube which projected through the cover of the crystal

[10] Nils G. Johnson, Dissertation (Lund, 1939), p. 50.
[11] H. M. O'Bryan, Rev. Sci. Inst. **5**, 125 (1934).

chamber, Fig. 2. Since a bent crystal requires a long focal spot, the cathode was a straight section of tungsten wire screened so as to produce a focal spot 18 mm long and 1 mm high. The aluminum foil, upon which the vapors were condensed to form the absorption film, was mounted so that it could be revolved about a vertical axis and stand either beside the mouth of the oven or in the path of the beam from the x-ray tube to the crystal. When the film was beside the mouth of the oven, it could be lowered over the oven by means of an arm attached to a ground metal joint mounted on the side of the crystal chamber. All other joints except the two for the manipulation of the film were flanged joints sealed by rubber-ring gaskets mounted in a groove in one of each pair of flanges.

The crystal holder and the cover carrying the x-ray tube could be rotated independently to give any glancing angle from 15° to 80°. The holder for the bent crystal was of the type designed by Haglund and described by Ohlin,[12] the opening in the front face being 1.5 mm by 20 mm.

The beam reflected by the crystal passed through a two-inch hole drilled through the cylindrical wall of the crystal chamber and through a cast-iron plate two inches thick and one foot square, to which the wall of the chamber was fastened. A shelf of cantilever construction was bolted to the other side of the cast-iron plate in order to support a track upon which the table bearing the plate holder moved. In order to protect the plate from visible light and scattered radiation, the plate holder was surrounded by an aluminum shield in which a window covered by two thicknesses of 0.6μ aluminum foil was provided for passage of the x-rays.

A flanged, steel pipe three feet long, ten inches in diameter, and blinded off at one end was mounted on a track to form the vacuum chamber for the plate holder when rolled forward to abut against the cast-iron plate. The degree of evacuation at any time could be determined by means of a Pirani tube and an ionization gauge, both permanently attached to the crystal chamber. The measuring circuit for these gauges was built according to the design of Sten v. Friesen.[13]

[12] Per Ohlin, Dissertation (Upsala, 1941), pp. 25, 26.
[13] Sten v. Friesen, Arkiv. f. Mat. Astronom. Fysik **B27**, No. 11 (1939).

Operation

The process of forming the absorption film was not started until the spectrograph and cathode-ray oven had been evacuated to 2×10^{-4} mm of Hg or less. To produce films of sodium metal a reaction mixture of equal parts by weight of powdered zirconium and sodium molybdate was employed.[14] The films of sodium halides were produced by evaporating C. P. halides. The evaporation process was always carried out by increasing the strength of the cathode-ray current in small steps, with sufficient time between successive steps to reduce each temporary rise in pressure. Whether or not a deposit had been evenly formed on the aluminum foil could be determined by tilting up the film holder until the under side of the foil could be viewed through a glass window in the side wall of the crystal chamber. The approximate thickness of the deposit was determined, after recording the absorption spectrum, by weighing the film and the aluminum foil.

A tungsten anticathode was employed as the source of the continuous radiation since it produced no interfering lines in the region of the sodium edge. Because the rectifying equipment available, four tubes in Greinacher coupling without condensers, did not prevent a half-sine-wave fluctuation of voltage, the effective voltage had to be maintained at or below 1.6 kv to prevent higher orders of radiation from obscuring

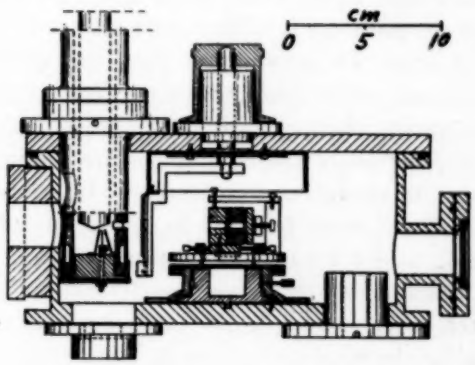

FIG. 2. Cross section of the crystal chamber.

[14] The use of this ratio was an oversight, since de Boer, Broos, and Emmens [Zeits. f. anorg. allgem. Chemie **191**, 113 (1930)] state that the product evaporated from a 1 to 1 mixture is sodium metal containing some oxide, while a mixture of 1 part sodium molybdate to 4 parts zirconium gives the pure metal.

the absorption spectrum at 1.1 kv. The $K\alpha_1$ line of sulfur always appeared on the plates, but not the $K\alpha_1$ line of potassium, which indicated that some second-order, but no third-order radiation was excited. Agfa Isochrome Plates, 18/10 Din, were employed for recording the spectra although their large grain size set a low limit to the enlargement which could be employed in making photometer curves.

Reference lines were exposed on one-half of the photographic plate either before or after the absorption spectrum was recorded. KCl and ZnO powders on a copper base on the other half of the anticathode were the source of the reference lines, 19-K $K\alpha_1$ at 3733.68 x.u. (effectively 11217.05 x.u. for a mica crystal in the third order) and 30-Zn $K\alpha_1$ at 1432.17 x.u. (11474.56 x.u. for mica in the eighth order).[15]

Measurements and Errors

A scratch was made on each plate parallel to the absorption edge, and the distances from this scratch to the reference lines were measured on a comparator. All measurements of positions of the edges and secondary absorption structures were made on photometer curves with respect to the sharp peak produced by this scratch. An abscissa enlargement of three was found suitable, the exact enlargement always being determined by recording at the same time a photometer curve of two scratches on a plate, the distances between the scratches having been determined by means of the comparator. A millimeter netting was also copied onto the photometer curve in order to provide automatic correction for shrinkage of the paper during photographic development.

The position of the edge and other main features on the photometer curves could be measured to ± 0.1 mm, corresponding to ± 0.03 mm on the plate or to ± 1.1 x.u. The positions of the reference lines could, of course, be determined with greater precision, while the precision was much lower for the weaker features of the spectra.

A number of systematic errors in wave-length measurements arise when a bent-crystal spectrograph is employed and the spectrum is recorded on a plane plate tangent to the focusing circle. Equations for calculating the various errors have been published by Sandstrom.[16] These errors arise from the geometry of the system even though the curved surface of the crystal is perfectly cylindrical. For the spectrograph employed in this work, the geometrical data required for Sandstrom's equations are: radius of the bent crystal 498 mm, crystal opening 20 mm $\times$ 1.5 mm, height of focal spot 1 mm, glancing angle 35° 49′, distance from focal spot to crystal 80 mm, and thickness of the photographic emulsion 12μ. Variations in temperature of the crystal are also a factor, and while the temperature of the crystal itself was not measured, the maximum temperature variation in the room during an exposure was 3°C.

At a crystal opening of 20 mm, which was that employed in obtaining most of the data, the largest error under these conditions is the focusing error, which arises because of the fact that the radius of curvature of the surface of the bent crystal is twice that of the focusing circle. When a crystal opening of 20 mm is employed, the focusing defect at a glancing angle of 35° should broaden a sharp line toward shorter wave-lengths into a beam 4.7 x.u. wide. A reduction of the crystal opening to 8 mm would reduce the focusing defect to 1.1 x.u., which should be practically counterbalanced by other effects so as to produce symmetrical broadening. Comparison of the positions of the reference lines when recorded with 20-mm and 8-mm openings revealed a shift toward shorter wave-lengths of only 0.4 x.u. at the larger opening.

In addition to errors owing to geometry of the system, the possibility of lack of definition owing to imperfect bending of the crystal and inaccurate focusing of the apparatus must be considered. The radius of curvature of the crystal was determined by the standard methods employed at the Institute of Physics at Upsala, which involve observation of the points of focus of x-ray emission lines on a nearly horizontally inclined photographic plate. From the radius of curvature so determined the required distance from the crystal to the plate can be calculated for any glancing angle. In order to check the sharpness of focus,

[15] Manne Siegbahn, *Spektroscopie der Rontgenstrahlen* (Julius Springer, Berlin, 1931).

[16] Arne Sandstrom, Zeits. f. Physik **84**, 541 (1933); **92**, 622 (1934); Dissertation (Upsala, 1935), pp. 21–27; also H. H. Johann, Zeits. f. Physik **69**, 185 (1931).

TABLE I. Absorption films, exposure times, and individual measurements of the main edges.

Substance	Absorption film mg/cm²	Film No.	Exposure Hours at 60 ma	Exposure Crystal opening mm	Plate No.	Wave-lengths for the main edges x.u. K′	K_1	K_2	K_3
Na	0.4*	24	12	20	24		11544.3		
	0.6*	35	16	20	35		11544.6		
	1.0*	37	27	20	37		11545.5		
	0.6*	39	12	20	39		11545.5		
Na_2O	0.6	35	5	20	36			11507.4	
	1.0	37	31	20	38		11523.8	11509.4	11488.8
	0.6	39	15	20	40		11520.1	11507.1	11480.6
		39	12	20	45		11522.1	11506.5	11484.5
NaF	0.5	46	26	20	46		11498.9		
		46	10	20	47	11517.5	11502.3		
		46	20	8	48	11518.5	11498.9		
	0.6	49	14	8	49	11519.8	11501.2		
		49	7	20	50	11519.0	11500.6		
		49	6	20	51	11522.0	11503.1		
		49	12	8	52	11523.7	11503.1		
NaCl	0.5	22	9	20	22		11500.3		
	0.9	28	18	20	28		11502.6		
	0.4	30	25	20	30		11501.2		
		30	38	20	31		11502.3		
NaBr	0.3	53	15	20	53		11504.6		
	0.5	54	8	20	54		11505.1		
		54	12	20	55		11504.6		
NaI	0.6	56	16	20	56		11507.1		
	0.2	59	22	20	59		11502.9		

* Weight given for oxidized film after removal from spectrograph.

estimates were made of the widths of emission lines exposed at large glancing angles with the full opening of the crystal. At a glancing angle of 58° the full width for 29-Cu $K\alpha_1$ (eleventh order) was found to be 0.6 x.u. at half-maximum intensity, which is comparable with the widths observed in a double-crystal spectrometer.[17] It is evident that even with the full crystal opening no significant broadening occurred such as might be expected to result from imperfect bending of the crystal.

It can be concluded that the systematic errors are smaller than the accidental errors in determining the wave-lengths. However, repetition of the measurements at higher dispersion might reveal additional structure in the edges, in which case some of the wave-lengths obtained might differ appreciably from those reported here.

TABLE II. Average wave-length and energy values for the sodium K-absorption edge.

Absorbing material	Edge	Wave-length x.u.	Wave-length Average deviation	Energy values ν/R	Energy values ev	Energy values Δ ev
Na	K_1	11545.0	0.6	78.932	1071.48	—
NaF	K'	11520.1	1.6	79.102	1073.79	2.32
	K_1	11501.2	1.1	79.233	1075.56	4.05
NaCl	K_1	11501.6	0.9	79.229	1075.52	4.03
NaBr	K_1	11504.8	0.3	79.206	1075.22	3.73
NaI	K_1	11505.8	2.1	79.206	1075.28	3.73
Na_2O	K_1	11522.6	1.3	79.090	1073.62	2.16
	K_2	11507.6	0.9	79.189	1074.96	3.49
	K_3	11484.6	2.3	79.347	1077.11	5.64

[17] Lyman G. Parratt, Phys. Rev. **50**, 1–15 (1936).

DATA

Table I lists each absorption film from which absorption spectra were obtained, together with the wave-length values found for each of the main edges. Table II gives the average value obtained for the wave-length of each edge and the corresponding average energy values, calculated from the relations $\nu/R=10^{11}/109737\lambda$ and $ev=12370.2\times10^3/\lambda$, where λ is in x units. Average wave-lengths for the edges and all reproducible features in the secondary absorption structures are listed in Table III. The figures in the third column of each compound in Table III indicate the average deviation of the individual measurements from the average, and the figures in the fourth column indicate the number of

TABLE III. Wave-length values for the features of the sodium K-absorption edge.

Feature	Av. λ x.u.	Av. dev.	No. of measurements	Feature	Av. λ x.u.	Av. dev.	No. of measurements
	NaF				NaCl		
K'	11520.1	1.6	(6)	K_1	11501.6	0.9	(4)
K_1	11501.2	1.5	(7)	A	11492.6	0.5	(4)
A	11481.2	2.9	(7)	α	11479.0	3.1	(4)
a	11447.3	3.5	(7)	B	11462.0	1.1	(4)
b	11435.8	5.3	(7)	a	11442.4	1.9	(3)
α	11426.4	1.0	(7)	β	11424.6	1.1	(4)
B	11399.1	3.3	(7)	C	11412.1	2.5	(4)
c	11371.2	0.9	(3)	γ	11402.1	1.5	(4)
d	11353.1	1.6	(3)	D	11383.5	1.2	(4)
C	11347.5	5.5	(4)	δ	11323.6	3.6	(4)
γ	11332.7	3.2	(4)	E	11286.3	4.3	(4)
D	11299.7	4.3	(7)	F	11216.9	6.0	(4)
δ	11223.6	3.2	(7)	ϵ	11164.5	9.0	(4)
E	11156.9	7.0	(7)	G	11094.5	5.2	(4)
F	11058.4	7.5	(6)	ζ	10917.4	1.8	(2)
ϵ	10974.7	9.1	(7)	H	10765.8	5.4	(2)
G	10869.1	8.7	(7)				
	NaBr				Na_2O*		
K_1	11504.8	0.3	(3)	K_1	11522.0	1.3	(3)
A	11495.4	0.5	(3)	a	11517.2	0.2	(2)
α	11486.7	1.5	(3)	K_2	11507.6	0.9	(4)
a	11479.6	1.9	(2)	b	11499.2	1.4	(2)
B	11468.9	0.8	(3)	K_3	11484.6	2.8	(3)
b	11452.3	0	(2)	A	11464.2	0.8	(3)
β	11430.0	1.8	(3)	α	11425.0	7.7	(2)
C	11421.8	1.1	(3)	B	11410.0	1.7	(2)
γ	11412.8	0.9	(3)	β	11247.7	7.4	(3)
D	11399.5	1.4	(3)	C	11093.7		(1)
δ	11335.8	5.3	(3)	γ	10932.8		(1)
E	11298.2	7.9	(3)				
F	11244.8	3.3	(2)				
ϵ	11198.5	9.7	(3)				
G	11032.7		(1)				
ζ	10945.8	17	(2)				
	NaI				Na		
K_1	11505.0	2.1	(2)	K_1	11545.0	0.6	(4)
A	11499.0	0.5	(2)	K_2	11516.4	7.4	(4)
α	11487.8	0.6	(2)	K_3	11489.4	0.8	(4)
a	11481.2	0.5	(2)	A	11439.6	13	(4)
B	11471.8		(1)	α	11266.4		(1)
b	11446.6		(1)	B	11114.7		(1)
β	11430.1		(1)	β	10958.2		(1)
D	11420.8		(1)	C	10677.6		(1)

* Film of metallic sodium oxidized by exposure to air.

measurements (number of plates) entering into the determination of each average. Typical photometer curves are reproduced in Figs. 3 and 4.

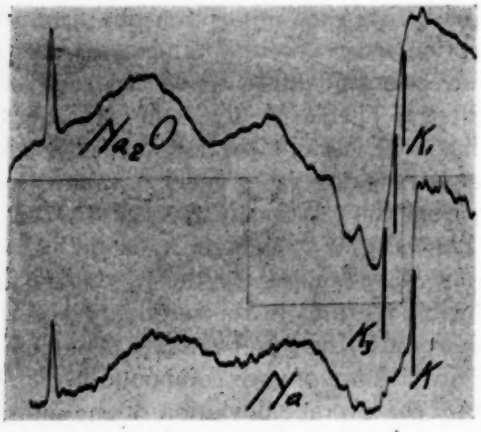

FIG. 3. Photometer curves of the edge for a film of sodium metal and for an air-oxidized film.

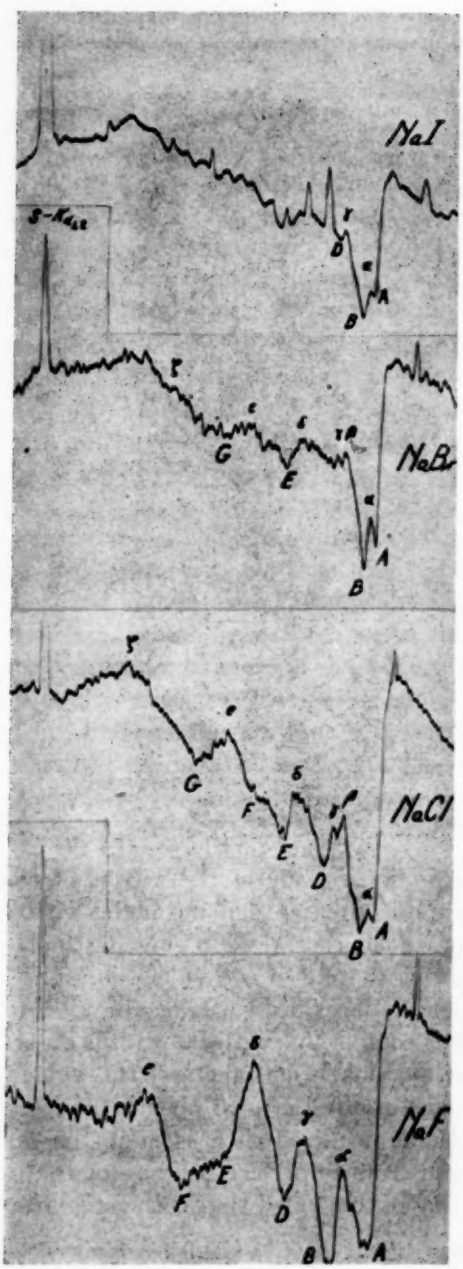

FIG. 4. Photometer curves of the edges for the sodium halides. Reading from top to bottom, the curves correspond to plates Nos. 59, 54, 28, and 49.

In all tables the main edge is distinguished by K_1, the minima on the photometer curves (i.e., maxima of absorption) are symbolized by Roman capitals, and the maxima on the photometer curves (minima of absorption) by Greek letters.

TABLE IV. The energy levels for sodium metal.*

Level		ev	ν/R
K	$1s$	1071.5	78.93_2
$L_{II\ III}$	$2p$	30.7	2.26_9
$M_{II\ III}$	$3p$	0.64	0.04_8

These symbols have been written in on the photometer curves. The small Roman letters in Table III designate less conspicuous features, whose positions are not indicated in Figs. 3 and 4, but which can perhaps be discerned in relation to the designated maxima and minima.

In connection with the discussion to follow it should be noted that only those features which are similar in appearance (width, strength, and shape) as well as in relative position have received like designations in the photometer curves of the spectra for the several halides. The secondary structures for the spectra of sodium metal and sodium oxide are quite distinct from those of the halides, however. The features α, B, β, and C in the spectrum for sodium metal and C and γ in the spectrum for sodium oxide appeared in each case on only one plate and may be fictitious.

DISCUSSION

(a) The Main Edges

The $K\beta_x$ emission line for Na and Mg has been considered in the past[18] as representing a semi-optical transition since this line is assigned to the transition $K-M_{II\ III}$, and there should be no electrons in the $M_{II\ III}$ shell for these elements. Wetterblad[19] plotted the wave-length differences between the $K\alpha_{12}$ and the $K\beta_x$ emission lines for the elements from 17-Cl to 11-Na and found that the curve changed sharply in slope at 12-Mg and 11-Na. A plot of the difference between the K edge and the $K\beta_x$ line in terms of ev, Fig. 5, gives a smooth curve from Si through Na, however. The energies of the K-absorption edges also agree within 1 or 2 ev with the energies of the $K\beta_x$ lines. Since the $K\beta_x$ emission actually constitutes a band, careful determination of the short wave-length edges of these bands for the elements from 17-Cl to 11-Na would undoubtedly result in a closer agreement with the K-absorption edges. It therefore appears unnecessary to consider the $K\beta_x$ line as a semi-optical transition for Na and Mg, but as a transition from a valence shell, which may be designated $M_{II\ III}$. Farineau[20] has previously reached a similar conclusion as to the origin of this line in the case of Mg, Al, Si. With this assignment for the $K\beta_x$ emission line, the wave-lengths for the K-absorption edge and the $K\beta_x$ emission line,[21] together with the value of 11.9093A = 11885.2 x.u. obtained by Tyrén[22] for the $K\alpha_1$ line, can be used to calculate the energies of the $1s$, $2p$, and $3p$ levels for metallic sodium. The energy values obtained are shown in Table IV.

The main edge for sodium metal is followed by two very weak edges, K_2 and K_3. Their wave-

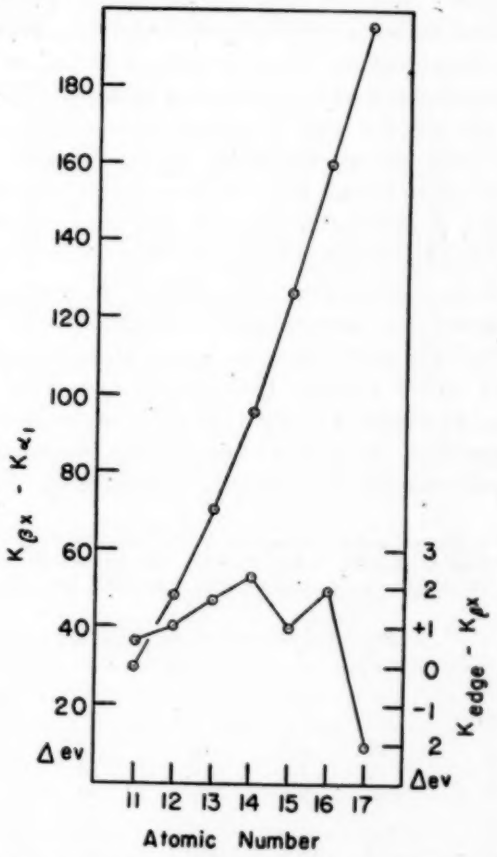

FIG. 5. Energy differences for $K\beta_x - K\alpha_1$ and K edge $- K\beta_x$, for the elements sodium to chlorine.

[18] E. Bäcklin, M. Siegbahn, R. Thoraeus, Phil. Mag. 49, 513 and 1326 (1925). Reference 15, p. 392.
[19] Thorsten Wetterblad, Zeits. f. Physik 42, 603 (1927).
[20] Jules Farineau, Comptes rendus 208, 1725 (1939).
[21] Reference 15, p. 165.
[22] Folke Tyrén, Nova Acta Reg. Soc. Sci. Upsala [IV], 12, 39 (1940).

TABLE V.

	K_1 absorption edge		$K\beta_1$ emission line[a]	
Compound	λ, x.u.	ev	λ, x.u.	ev
NaF	11501.2	1075.6	11637	1063.0
NaCl	11501.6	1075.5	11633	1063.4
NaI	11504.8	1075.2	11621	1064.5
NaBr	11505.0	1075.2	11614	1065.1

[a] See reference 19.

lengths correspond approximately to the average of the wave-lengths of the K_1 and K_2 edges and to the wave-length of the K_3 edge of the oxide, respectively. As mentioned in the experimental section, there is reason to believe that the sodium film contained some oxide.

The main edge for the oxide lies 23 x.u. toward shorter wave-length, or 2.1_5 ev toward higher energies, than the edge for the metal. Furthermore, it no longer coincides with the $K\beta_1$ emission line for the oxide, which is shifted 43 x.u. in the opposite direction as compared with the $K\beta_x$ line for the metal.[21] The directions of these shifts are the same as are found in analogous cases for elements in higher series of the periodic system.[23]

The K-absorption edge for Na^+ in its halides is shifted still farther than for the oxide toward shorter wave-lengths. The shift is some 40 x.u., or about 4 ev toward higher energies. From NaF to NaI the shift decreases from 4.0_9 to 3.7_3 ev, a shift which exceeds the probable error of the data. In terms of energy the shifts are somewhat larger than the corresponding shifts observed by Lindh[24] for the K-absorption edge for K^+ in its halides. Our result for the K edge for Na^+ in NaCl can be compared with the value obtained by Johnson[10] in the course of a study of the effects of chemical combination on emission spectra. He found a wave-length of 11498 x.u., compared with the present result of 11501.6 x.u. In terms of energy the difference is 0.4 ev. Since Johnson's measurement was not intended to be a precision determination, the agreement between the two results is satisfactory.

In Table V the wave-lengths of the absorption edges for the halides are compared with Wetterblad's measurements of the $K\beta_1$ line for Na^+ in its halides. The energy for the $K\beta_1$ emission line for Na^+ in NaCl is 12.1 ev less than the energy for the absorption edge. Valasek[25] has plotted the separations in units of ev between the various K-emission lines and the K-absorption edge for Cl^- in NaCl. He predicted that if the difference between the $K\beta_1$ line and the K-absorption edge for Na^+ in NaCl were plotted in the same figure, the $K\beta_1$ line for Na^+ would coincide with the $K\beta_1$ line for Cl^-, which lies 7.1 ev from the edge. The difference of 5.0 ev between the predicted and observed positions for the $K\beta_1$ line for Na^+ is fully an order of magnitude greater than the probable experimental errors. It seems necessary to conclude either that the $K\beta_1$ line for Na^+ in NaCl does not represent the $Na^+-1s-Cl^--3p$ transition as predicted, or that the end levels for the K-absorption transitions for Na^+ and Cl^- do not both begin at the Cl^- continuum as assumed. Different end levels might be attributed to different selection rules.[7]

[23] V. Hugo Sanner, Dissertation (Upsala, 1941), p. 93.
[24] Axel E. Lindh, Arkiv. f. Mat. Astronom. Fysik **18**, No. 14 (1924). See also reference 15, pp. 289–290.
[25] Joseph Valasek, Phys. Rev. **53**, 274 (1938).

TABLE VI. Distance of the secondary-structure features to the main edge, in x.u., ev, and a^2 (ev).

	NaF			NaCl			NaBr			NaI		
Structure	x.u.	ev	a^2 ev $\cdot 10^{16}$	x.u.	ev	a^2 ev $\cdot 10^{16}$	x.u.	ev	a^2 ev $\cdot 10^{16}$	x.u.	ev	a^2 ev $\cdot 10^{16}$
A	20	1.9	40	8	0.9	29	10	0.9	32	6	0.6	29
α	75	7.0	150	23	2.2	70	18	1.7	60	17	1.6	67
B	102	9.6	205	40	3.8	120	36	3.4	121	33	3.1	129
β				77	7.3	231	75	7.1	252			
γ	168	16.0	326	100	9.4	298	92	8.7	309	75	7.0	292
D	201	19.1	408	118	11.2	354	105	10.0	355	84	7.9	330
δ	277	26.5	566	178	16.9	535	169	16.0	568			
E	344	33.1	706	215	20.5	649	207	19.7	700			
F	443	42.9	915	285	27.2	861	260	24.8	881			
ε	526	51.4	1097	338	32.4	1025	307	29.4	1044			
G	632	62.4	1331	408	39.4	1248	472	45.9	1630			
ζ				585	57.4	1817	559	54.8	1948			
H				735	73.3	2320						

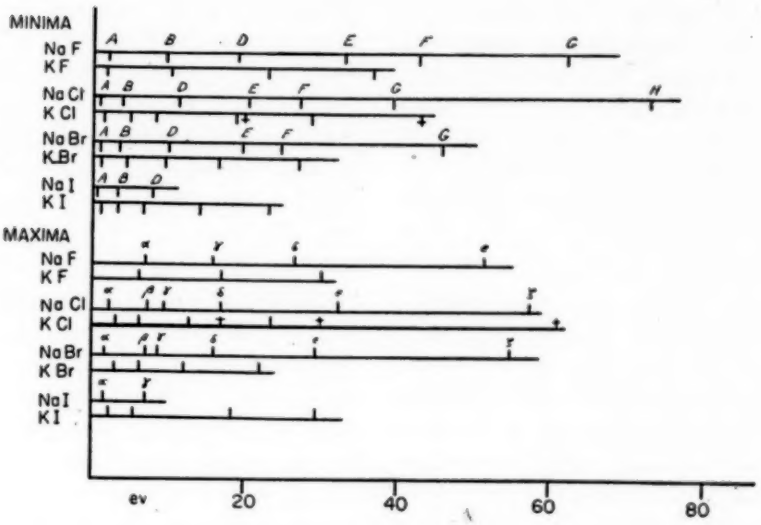

FIG. 6. Separation in electron volts between the secondary structure and the main edges for Na^+ and K^+. + indicates data from Coster and Klamer (reference 7). Other data for K halides taken from Brewington (reference 8).

In the case of KCl, also, the coincidence between the $K\beta_5$ line for K^+ and the $K\beta_1$ line for Cl^- is not perfect, the $K\beta_5$ line for K^+ lying 1.2 ev farther out from the edge. Valasek attributed this discrepancy, however, to an error in Stelling's energy value[26] for the K-absorption edge for Cl^- in KCl.

(b) Secondary Absorption Structure for the Halides

As was mentioned in the introductory section, in view of the difficulties of carrying out detailed calculations of the shape of the secondary absorption pattern which should arise from the overlapping of the numerous forbidden energy zones predicted by Kronig's theory, the theory has usually been indirectly tested in case of ionic lattices by comparison of the secondary structures for a particular absorption edge for a series of similar lattices of varying lattice distances. For a cubic type of lattice the separation of the structure from the edge in terms of energy should be inversely proportional to the square of the lattice distance. Brewington[8] applied this criterion in the case of the secondary absorption patterns accompanying the K-absorption edges for K^+ and Cl^- in the series of potassium halides and alkali chlorides. He came to the conclusion that his data did not agree with the hypothesis. The same criterion will now be applied to the data given in Table III, and since the potassium and sodium halides belong to the same crystal type and the K-absorption edges for sodium and potassium should represent the same transition, the data in Table III will be compared directly with the available data for the K edge of K^+ in the potassium halides.

The separation of each feature of the secondary structure from the main edge has been calculated from Table III and is listed in Table VI for the absorption spectrum from each halide. The separations are expressed in x.u., electron volts, and in electron volts multiplied by the square of the lattice constant of the halide. The values employed for the squares of the lattice constants are 21.34, 31.65, 35.52, and 41.73 cm$\times 10^{-16}$ for NaF, NaCl, NaBr, and NaI, respectively. In Fig. 6, the separations shown in Table VI in terms of electron volts are plotted together with similar data of Brewington[8] and of Coster and Klamer[7] for the secondary absorption structure accompanying the K-absorption edge for potassium in its halides. This comparison does not reveal very striking likenesses between the spectra for the different halides, except for the features close to the main edges. In Fig. 7, however, in which the

[26] Reference 15, p. 282.

abscissas represent separations multiplied by the squares of the lattice constants, the similarities are more marked. It is true that the positions of the first two maxima and minima for the other halides are not matched in the spectra for NaF and KF. On the other hand, considering the spectra for the sodium halides alone, the double minimum E-F is exceptionally strong for NaF, obvious for NaCl, and still apparent, though very weak, for NaBr, and they lie on a nearly vertical line in Fig. 7. Aside from the discrepancies mentioned for NaF and KF, even the differences between the patterns for the sodium and potassium halides are not very striking in view of the facts that the spectra were recorded in different laboratories and are based on measurements at two different wave-lengths in which the dispersions obtainable, in terms of energy units, differ greatly. With reservations, it can therefore be said that the pattern in Fig. 7 confirms the inverse proportionality with the squares of the lattice distances, as required by Kronig's hypothesis.

On the other hand, Brewington's data for the absorption structures accompanying the *K*-absorption edge for Cl^- in the various alkali chlorides show very significant variations from the relations predicted by Kronig's hypothesis. It is conceivable that there may be some principal difference between the secondary structure accompanying absorption by anions as compared with cations, such as different selection rules for the electronic transitions.[7] The fact, discussed in the foregoing section, that the energy difference between presumably similar emission and absorption transitions is not the same for Na^+ and Cl^- transitions in NaCl, or for K^+ and Cl^- transitions in KCl, is in harmony with the supposition that electron transitions involving anions in ionic lattices obey different selection rules from those involving cations. Additional experimental data concerned with the secondary absorption structure accompanying absorption edges of anions would be desirable. It is suggested that a study of the secondary absorption patterns accompanying the *K*-absorption edge for fluorine in the series of alkali fluorides, whose lattice constants extend over a wide range, would serve to make any differences very evident when compared with the present patterns for the sodium halides.

ACKNOWLEDGMENTS

The author is pleased to acknowledge his deep indebtedness to Professor Axel H. Lindh, Director of the Institute of Physics of the University of Upsala, who, under the difficult conditions prevailing in a war-threatened country,

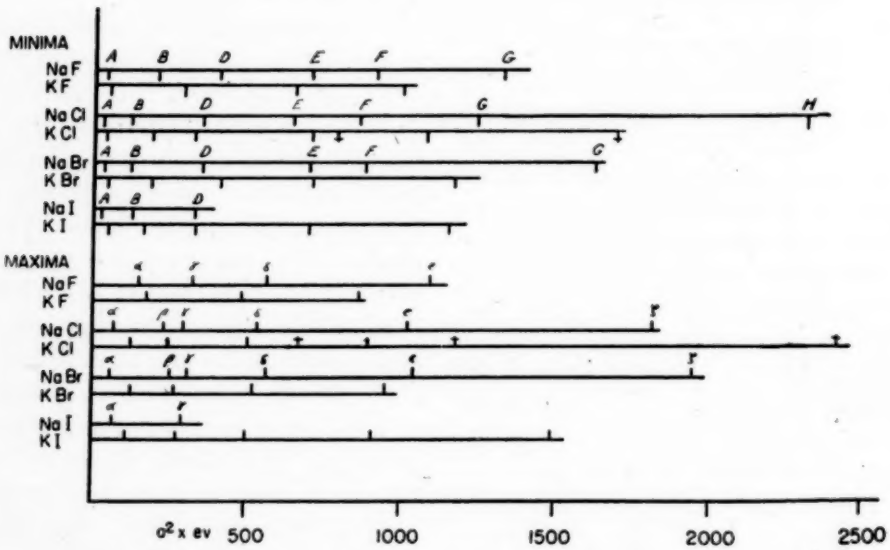

FIG. 7. Separation between the secondary structure and the main edges in a^2 ev $\cdot 10^{16}$. + indicates data from Coster and Klamer (reference 7). Other data for K halides taken from Brewington (reference 8).

put material, skilled machinists, and a laboratory at the author's disposal during 1940 and 1941. He is also indebted to Mr. John Amberntson for skilled workmanship, to many Fellows of the Institute, including Åke Nilsson, Philip Haglund, Per Ohlin, and Harald Flemberg for advice, and to Professor Manne Siegbahn and Torsten Magnusson for an introduction to the experimental techniques of x-ray spectroscopy. For any shortcomings in conception or execution of the work or in interpretation of results, the author has sole responsibility. Finally, acknowledgment is made of financial assistance from the American Scandinavian Foundation.

PHYSICAL REVIEW VOLUME 66, NUMBERS 7 AND 8 OCTOBER 1 AND 15, 1944

K-Electron Capture in Radioactive Argon A^{37}

PAUL K. WEIMER, J. D. KURBATOV, AND M. L. POOL
Department of Physics, The Ohio State University, Columbus, Ohio.

(Received May 14, 1942)

A radioactive gas possessing a half-life of 34.1±0.3 days has been produced by bombarding solid samples containing potassium, chlorine, calcium, or sulfur. The observed data require that this activity be assigned to A^{37} in accord with the following reactions:

$$K^{39}+H^{2}\rightarrow A^{37}+He^{4}$$
$$Cl^{37}+H^{2}\rightarrow A^{37}+2n$$
$$Cl^{37}+H^{1}\rightarrow A^{37}+n$$
$$S^{34}+He^{4}\rightarrow A^{37}+n$$
$$Ca^{40}+n\rightarrow A^{37}+He^{4}.$$

Alpha-bombardment of chlorine has failed to produce a measurable quantity of this activity by an (α, d) reaction. The detection of 4.72A Cl $K\alpha$ x-rays by means of critical absorption measurements in gaseous compounds of chlorine and sulfur reveals that A^{37} decays by *K*-electron capture. Cloud-chamber observations and aluminum absorption measurements indicate that *K*-electron capture accounts for more than 99.9 percent of the total number of disintegrations. The probability of the x-ray quantum being internally converted has been calculated to be 0.96±0.03. No gamma-rays were observed.

INTRODUCTION

THE first radioactive argon to be reported[1] was produced by the bombardment of the gas itself with deuterons. This activity, which decayed with a half-life of 1.83 hours, was assigned to the isotope A^{41}. The radiation emitted was found to consist of beta-rays[2] with a maximum energy of 1.5 Mev and of a gamma-ray[3] of 1.37 Mev. More recently, a 1.9-sec. radioactive argon A^{35} has resulted from the bombardment of chlorine with protons[4] and of sulfur with alpha-particles.[5] This isotope was found to emit positrons with a maximum energy of 4.38 Mev.

The stable isotopes of argon have the mass numbers 36, 38, and 40. The existence of isotopes A^{37} and A^{39} had never been proved although it is evident from the nuclear chart, shown in Fig. 1, that they might be produced by the bombardment of solid samples containing one of the neighboring elements. However, if an isotope of argon formed in this manner should decay entirely by *K* capture, then the resulting $K\alpha$ x-rays of chlorine (4.72A) would be largely absorbed in the sample itself. It is well known that a sensitive means for detecting a soft radiation emitted by a radioactive gas is obtained by admitting the

[1] A. H. Snell, Phys. Rev. **49**, 555 (1936).
[2] F. N. D. Kurie, J. R. Richardson, and H. C. Paxton, Phys. Rev. **49**, 368 (1936).
[3] J. R. Richardson and F. N. D. Kurie, Phys. Rev. **50**, 999 (1936).
[4] M. G. White, E. C. Creutz, L. A. Delsasso, and R. R. Wilson, Phys. Rev. **59**, 63 (1941).
[5] L. D. P. King and D. R. Elliott, Phys. Rev. **59**, 108A (1941).

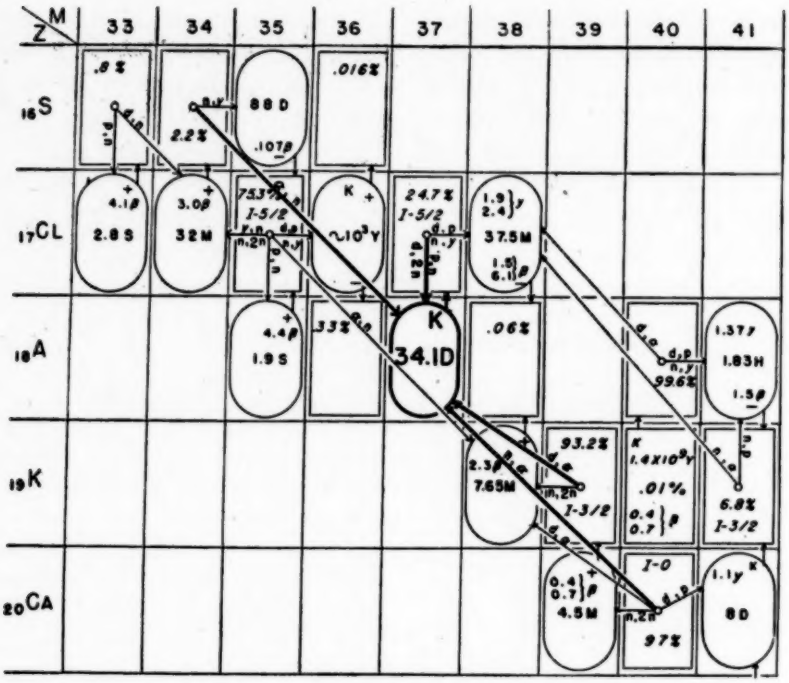

FIG. 1. Nuclear chart in the region of argon.

gas directly into an ionization chamber. It seemed of interest, therefore, to perform experiments in which the radioactive argon is extracted from the bombarded sample and conducted into an ionization chamber.

The 34.1-day radioactive gas which was found by this procedure has already been reported[6] briefly and has been assigned to the isotope A^{37}. A thorough study of the radiation emitted by this isotope has now been made, and the results of this investigation are presented here, along with a more detailed description of the previous experiments leading to the isotopic assignment.

TABLE I. Relative yields from various reactions.

Bombardment	Yield	Nuclear reaction
KCl+H^2	540	$K^{39}(d, \alpha)A^{37}$ $Cl^{37}(d, 2n)A^{37}$
KF+H^2	440	$K^{39}(d, \alpha)A^{37}$
NaCl+H^2	120	$Cl^{37}(d, 2n)A^{37}$
NaCl+H^1	97	$Cl^{37}(p, n)A^{37}$
FeS_2+He^4	4.6	$S^{34}(\alpha, n)A^{37}$
CaO+n	1	$Ca^{40}(n, \alpha)A^{37}$
NaCl+He^4	0	$Cl^{35}(\alpha, d)A^{37}$

[6] P. K. Weimer, J. D. Kurbatov, and M. L. Pool, Phys. Rev. **60**, 469 (1941).

ISOTOPIC ASSIGNMENT AND DECAY PERIOD

Bombardments were made in the 42-inch cyclotron of The Ohio State University. A small quantity of the substance to be bombarded was placed in a copper target holder which was inserted between the dees. A 2.40 mg/cm² aluminum foil was placed in front of the sample to lessen contamination, but no attempt was made to seal the target to prevent the escape of the radioactive gas formed as a disintegration product.

The gas was extracted from the bombarded sample by dissolving the sample, then bubbling air through the solution and allowing it to pass into an ionization chamber which had been previously evacuated. Suitable precautions were taken to prevent volatile radioactive substances other than rare gases from entering the chamber. The ionization chamber, which had a volume of about two liters, was connected to a Wulf bifilar electrometer.

A summary of bombardments with the relative yield of each is given in Table I. Since the activity was obtained from a deuteron bombardment of potassium, the active isotope might be either A^{37},

A^{39}, or possibly A^{38}. As the 34.1-day activity was obtained by bombardment of chlorine with protons, the activity must therefore be assigned to the isotope A^{37}. The remaining bombardments listed in Table I are all consistent with this assignment. Further evidence that the 34.1-day activity is not owing to A^{39} was given by a neutron bombardment of potassium which produced only the 1.83-hour A^{41}.

The decay curves for each bombardment are shown in Fig. 2. The half-life was observed to be 34.1 ± 0.3 days with no indication that any other period was produced by any of the bombardments.

PROOF OF *K*-ELECTRON CAPTURE

Critical Absorption Measurements

If A^{37} decays by *K*-electron capture, the characteristic 4.72A chlorine $K\alpha$ x-rays would be emitted. This wave-length may be identified by the fact that it will be very strongly absorbed in sulfur but only slightly absorbed in chlorine. Because of the high absorption coefficients involved, it was considered most convenient to use the gases hydrogen sulfide and freon (CCl_2F_2) as absorbers.

For this purpose a special ionization chamber, shown in Fig. 3, was constructed. The radioactive argon was confined at atmospheric pressure in the small chamber marked *A*. A thin aluminum window of 2.4 mg/cm² thickness separated the active gas from the absorbing gas

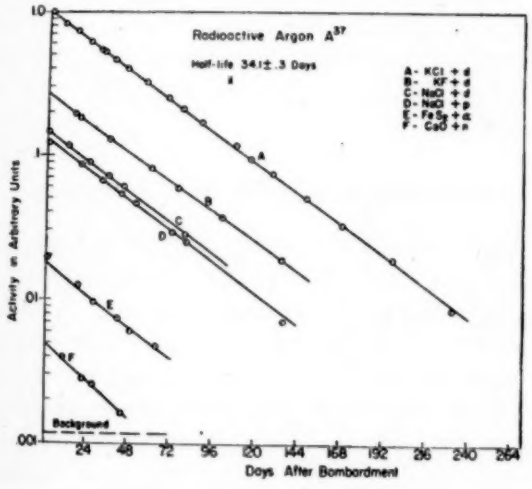

FIG. 2. Decay curves for argon A^{37}.

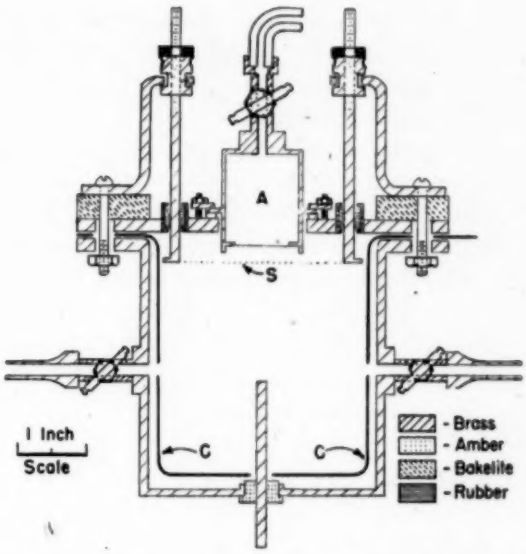

FIG. 3. Gas absorption ionization chamber, showing movable screen *S*, inner cylinder *C*, and position of small gas chamber *A*.

which was placed at atmospheric pressure in the ionization chamber. The inner cylinder *C* and the movable copper screen *S* were kept at 135 volts negative with respect to chamber *A* and the central collecting rod. Hence, only the ions formed below the screen had their charge collected by the central rod and measured by the electrometer. Thus by adjusting the distance of the screen from the foil, the thickness of the absorbing gas could be varied.

The advantage of this method is that it is unnecessary to vary the pressure of the absorbing gas as is done in the usual methods of measuring gas absorption. Therefore, a much thinner foil could be used for separating the active gas from the absorbing gas than could be used otherwise. This advantage is very great since an aluminum foil with a thickness of only 2.40 mg/cm² will reduce the intensity of a 4.72A x-ray to 8 percent of its initial intensity.

The collecting rod of the gas absorption chamber was connected to a sensitive Wulf unifilar electrometer. A 12-hour deuteron bombardment of KCl was found necessary to produce sufficient argon to give a measurable activity through the aluminum window. Even then, the activity was weak and amounted to only 1.5 times background when the screen was closest to the foil.

In Fig. 4, the ionization current obtained when chamber *A* of Fig. 3 was filled with argon A^{37} is plotted against the corresponding screen position. The more rapid initial drop of the ionization current as the screen was lowered in hydrogen sulfide (curve *C*) than was observed in the case of freon (curve *B*) indicates that the absorption coefficient in hydrogen sulfide is the greater in spite of the fact that its density is less. However, when an additional 1.7 mg/cm² (0.25 mil) aluminum foil was placed over the aluminum window of chamber *A*, curves *D* and *E* show identical absorption in freon and in hydrogen sulfide.

The decrease in ionization current as the screen is lowered is not entirely owing to absorption in the gas since at the same time the effective part of the ionization chamber is being decreased in volume and is being moved a greater distance from the radioactive source. The effect of these factors alone may be studied by filling chamber *A* with a radioactive gas emitting a much harder radiation. A radioactive gas which has been found to be convenient for this purpose is the 1.44-day krypton $Kr^{79,81}$ which emits beta-rays[7] with maximum energy of 0.4 Mev. With chamber

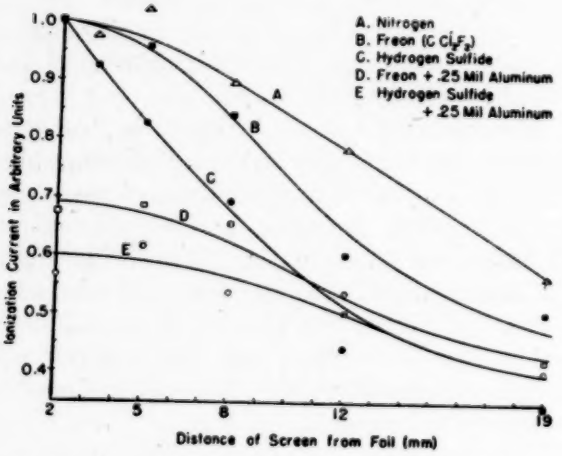

FIG. 4. Gas absorption curves for argon A^{37}, showing ionization current as a function of the position of the movable screen *S*.

A containing krypton, the ionization currents were measured when the screen was lowered in freon, hydrogen sulfide, and in air, and curves were plotted in a similar manner to those shown for argon in Fig. 4. In the case of krypton, the ionization current decreased less rapidly as the screen was lowered, and the curves for freon, hydrogen sulfide, and air were identical within the limits of experimental error.

Since a 1.7 mg/cm² aluminum foil should reduce an x-ray beam of 4.72A to 16 percent of its initial intensity, it is apparent from Fig. 4 that in addition to the x-rays, there is a harder component present in the radiation which passes through the aluminum window. Hence, by subtracting the ordinates of curves *D* from *B*, and *E* from *C*, the approximate intensity of the x-rays alone may be calculated at each screen setting. After making a correction for the dependence of the sensitivity of the ionization chamber upon the screen position, the experimental adsorption curves for the x-rays in freon and in hydrogen sulfide were plotted on semilog graph paper. These curves were in satisfactory agreement with the calculated absorption curves for 4.72A x-rays in these gases, with the absorption in hydrogen sulfide greater than that in freon by about the expected amount. Only x-rays lying between the critical absorption limits 4.38A and 5.008A will possess absorption curves in these gases which compare even approximately with the observed absorption curves. If the soft component of the radiation passing through the foil were owing to very low energy electrons, then a larger absorption coefficient should have been observed in the more dense freon.

It is therefore concluded from the critical absorption measurements that the characteristic $K\alpha$ x-rays of chlorine have been identified and that *K* capture in radioactive argon A^{37} is confirmed.

Aluminum Absorption Measurements

Aluminum absorption measurements on the radiation emitted by the argon were taken after the gas absorption experiments indicated the presence of an additional harder component passing through the foil. The absorption curve obtained from these measurements is shown in Fig. 5. The data for the points indicated by open circles were taken with the gas absorption ionization chamber by placing additional aluminum

[7] E. C. Creutz, L. A. Delsasso, R. B. Sutton, M. G. White, and W. H. Barkas, Phys. Rev. **58**, 481 (1940).

foils below the window. The data for the solid circles were obtained by mounting chamber *A* above a more sensitive ionization chamber filled with freon at 20-lb. pressure. The resulting curve shows clearly that there are two components

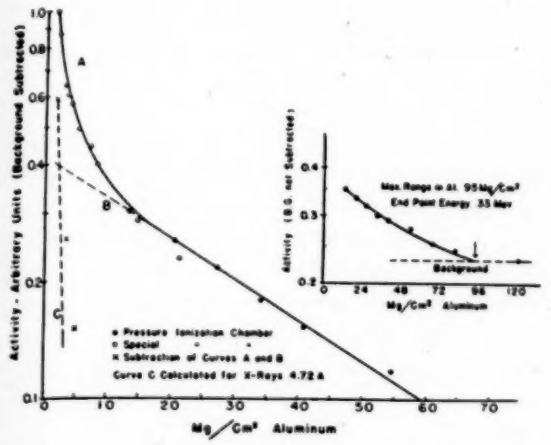

FIG. 5. Aluminum absorption curve for argon A^{37}. Section of curve labelled *A* shows effect of x-ray component, while *B* is owing to the harder component of range 95 mg/cm² in aluminum. Curve *C* represents the absorption in aluminum of 4.72A Cl $K\alpha$ x-rays.

present in the radiation passing through the window. The points obtained after subtracting the harder component give a curve consistent with the calculated absorption curve for 4.72A x-rays in aluminum.

The range in aluminum of the harder component was observed to be 95 mg/cm². Since the intensity of the radiation was weak, it was not possible to determine whether this component was composed of positrons or electrons. However, if we assume that the component does consist of one or the other, then the range in aluminum is found to correspond[8] to an energy of 0.33 Mev. Because of the high specific ionization of the beta-rays and the high internal conversion coefficient of the x-rays as well as their large absorption in the aluminum window, it follows that the harder component must represent a very small fraction of the total disintegrations occurring inside chamber *A*. The results of Table II, discussed in a later section, can be used to calculate this fraction, and it is found that regardless of whether the harder component is owing to positrons from A^{37} or to a beta-emitting impurity, the component cannot account for more than 0.1 percent of the total disintegrations.

[8] P. Curie, *Radioactivité* (Hermann and Company, Paris, 1935), p. 294.

Cloud-Chamber Observations

A small quantity of radioactive argon mixed with air and hydrogen was admitted into a cloud chamber. A typical photograph, shown in Fig. 6, reveals many spots and clusters of drops but no electron tracks. Similar spots appearing in cloud-chamber photographs have been interpreted by others[9] as owing to short range photoelectrons ejected by soft x-rays. Such spots or clusters probably represent single disintegrations. Clusters may be formed instead of spots when the time interval between the disintegration and the expansion is sufficient for diffusion of the ions.

When A^{37} decays by *K*-electron capture, an x-ray photon of 2600 electron volts is emitted by the atom provided the photon is not internally converted in an outer electron shell. In case conversion does take place, an Auger electron will be ejected from the atom with an energy ranging from 2400 to 2600 volts. An electron of this

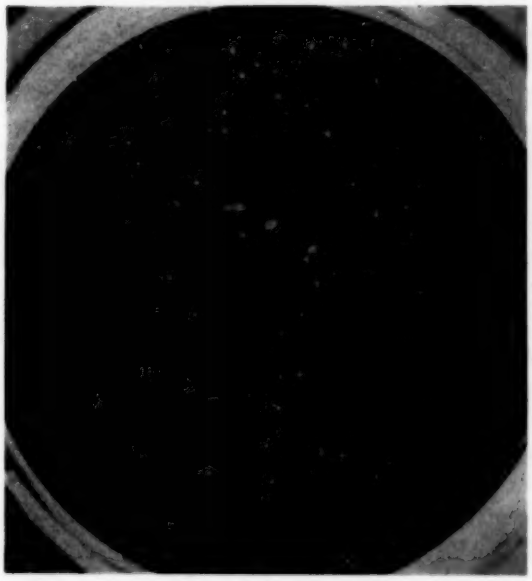

FIG. 6. Cloud-chamber photograph of argon A^{37} obtained by admitting the gas into expansion chamber. Spots and clusters of drops are due to Auger electrons and short range photoelectrons.

[9] E. J. Williams and E. Pickup, Nature **141**, 199 (1938).

energy has a range in air of about 0.04 cm. Thus a spot or cluster may be produced by an Auger electron emitted by the disintegrating atom itself as well as by a photoelectron ejected from a neighboring atom by the x-ray.

INTERNAL CONVERSION COEFFICIENT OF X-RAYS FROM *K* CAPTURE

When radioactive argon mixed with air is admitted directly into an ionization chamber, each disintegration by *K* capture releases 2600 ev of energy regardless of whether the x-ray is internally converted or not. However, when the gas is enclosed in chamber *A* and the ionization is owing to the radiation penetrating the foil, only those atoms in which the x-ray quantum is not internally converted are contributing to the ionization current. The Auger electrons, having an energy of less than 2600 ev, are entirely too weak to penetrate the 2.4 mg/cm² aluminum window. It is then apparent that a possibility exists for calculating the internal conversion coefficient for the x-rays by observing the ionization current produced through the foil by the argon, and then admitting the same gas directly into an ionization chamber. When the argon is in chamber *A* it is, of course, necessary to correct for the absorption of the x-rays in the aluminum window and for the less advantageous geometry. In order to make a better estimate of the geometrical factor, the measurements for argon have been compared with those for krypton Kr[70, 81].

In Table II are recorded the activities of argon and krypton each measured first in chamber *A* and then directly inside an ionization chamber. The value of *R* represents the factor by which the ionization current is multiplied when the gas is placed inside the ionization chamber. In the case of krypton, where the radiation consists of 0.4-Mev positrons, the absorption in the aluminum window may be considered negligible, and the value of *R* is determined largely by geometrical conditions. However, for argon *R* is influenced also by the absorption of the x-rays in the aluminum window and by the internal conversion of the x-rays. The probability that the x-ray quantum will not be converted equals approximately $(34.2)(0.60)/(6910)(0.08)=0.037$. The quantity (0.60) appears because only 60 percent of the radiation observed through the window of chamber *A* is composed of x-rays while the factor (0.08) corrects for the x-ray absorption in the aluminum window. Therefore, the value of the internal conversion coefficient which represents the probability of conversion is found to be 0.96 ± 0.03.

TABLE II. Relative activities of A and Kr.

		Argon	Krypton
T	Half-life (days)	34.1	1.44
t	Bombardment time (hours)	12	2
I_1	Activity of gas in chamber *A*	0.060	1.13
I_2	Activity of same gas in ionization chamber	415	38.6
R	Ratio of I_2 to I_1	6910	34.2

The internal conversion coefficient of the $K\alpha$ x-rays of chlorine has been previously determined by means of fluorescent yield experiments in which the *K* electron is removed by irradiating the atom with x-rays. Although it may be questionable whether the internal conversion coefficient is the same when the *K* electron is captured by the nucleus as when the electron is ejected photoelectrically from the atom, a comparison of the values calculated by the two methods is of interest. An average value of the coefficients obtained by several observers[10] using the fluorescent yield method is 0.92, which is somewhat lower than the coefficient 0.96 ± 0.03 arrived at in this paper. A value also greater than 0.92 was calculated from theoretical considerations by Burhop.[11]

It is a pleasure to acknowledge the support received from the University Development Fund and from Mr. Julius F. Stone. Valuable assistance received through the graduate school is also acknowledged.

[10] A. H. Compton and S. K. Allison, *X-Rays in Theory and Experiment* (D. Van Nostrand Company, Inc., New York, 1935), p. 477.
[11] E. H. S. Burhop, Proc. Roy. Soc. **A148**, 272 (1935).

PHYSICAL REVIEW VOLUME 66, NUMBERS 7 AND 8 OCTOBER 1 AND 15, 1944

Probability Method Applied to the Analysis of Recrystallization Data

CECIL G. DUNN
General Electric Company, Pittsfield, Massachusetts
(Received July 12, 1944)

Probability equations are developed for the purpose of analyzing recrystallization data to determine what mechanisms may operate during recrystallization processes. The equations allow for inexact variations in orientation changes provided the variations are small. $P(E)=1.43\alpha^3$ gives the probability that a nucleation will result in an orientation falling within α radians of a specified one A on a chance basis. Other equations are given for the probability of occurrence of nearly common orientations for various types of groupings. Values of probabilities are given in tables.

INTRODUCTION

NEW crystals are produced whenever a single crystal is deformed a small amount and heat-treated above its recrystallization temperature. In general, the changes in orientation that occur cannot be described in terms of simple mechanisms such as twinning or as the rotation of the lattice about a single axis. Furthermore, if a mechanism is postulated that involves a large number of transformations, it may be of little use in predicting changes in the crystal lattice unless appropriate selection rules are also introduced to limit the number of final orientations. Such rules would depend upon experimental conditions.

In an attempt to overcome some of these difficulties, the writer has found it useful to analyze recrystallization data with the aid of probability equations. These equations are developed in the present paper. The first relationship, Eq. (1), gives the probability that a single nucleation will produce an orientation falling within α radians of a specified one A assuming all possible orientations as equally likely. Any nucleation that satisfies this condition is called event E and its probability, $P(E)$. The reciprocal of $P(E)$ will be called N. It is the maximum number of events similar to E. We might say that there are N ways for an orientation to occur and if each way is equally likely, then the probability that it is E is $1/N$. Call each of the N configurations a state (a classified orientation). These states, instead of orientations, may be used in considering mechanisms of recrystallization. The number of states allowed for a given mechanism, however, must be considerably less than N if the mechanism is to have much meaning. In this connection, an illustration is given for a mechanism that would produce high order twins. Finally, an application is made of $P(E)$ to obtain Eqs. (2a) and (2b) for the probability of occurrence of new crystals having nearly common orientations.

PROBABILITY $P(E)$ OF OCCURRENCE OF AN ORIENTATION IN A GIVEN STATE

Consider all orientations that agree within α radians of A, defined by three mutually perpendicular poles A_1, A_2, A_3 of {100} planes, as belonging to the set E, i.e., in a state E. An orientation B belongs to E if no {100} pole of B departs from A by more than α radians. We ask, "What is the probability of occurrence of event E if an orientation is selected at random from a group of random orientations?" If E is to occur, B must be taken from a subset of the group of random orientations having one pole within α radians of one pole of A. Consider the subset for pole A_1 and call a selection from it event F. Complete a set F by taking all such orientations. From this set make further selections, each having a pole within α radians of A_2, and thus obtain a set G. Finally, all orientations of this group that have a pole within α radians of A_3 may be taken to form a set H.

The probability of E (event E) will be $P(E)=P(F)P(G)P(H)$ where $P(F)$ is the probability of F, $P(G)$ is the probability of G after F has occurred, and $P(H)$ is the probability of H after F and G have occurred. [1,2a]

[1] Arne Fisher, *Mathematical Theory of Probabilities*, (The MacMillan Company, New York, 1922), p. 29.
[2] Thornton C. Fry, *Probability and Its Engineering Uses* (D. Van Nostrand Company, Inc., New York, 1928), (a) p. 116; (b) pp. 19, 20.

In order to obtain $P(E)$, the original group of random orientations will be represented by a plot of {100} poles on the surface of a hemisphere of radius R. Let k be the number of poles per unit area, thus giving a total of $2\pi R^2k$ poles and $2\pi R^2k/3$ orientations. Computing the area near A_1, we find that there are $2\pi(1-\cos\alpha)R^2k$ poles within α radians of A_1. Each pole belongs to a unique orientation. Consequently, there are $2\pi(1-\cos\alpha)R^2k$ orientations that belong to the set F. Expressing probability in the usual way, we get at once for low values of α:

$$P(F)=\frac{2\pi(1-\cos\alpha)R^2k}{2\pi R^2k/3}=3(1-\cos\alpha)\doteq\tfrac{3}{2}\alpha^2,$$

the error being only $\frac{3}{10}$ of one percent for α equal to 10°.

Figure 1 is a stereographic projection of the plot of set F. The poles have a density k in the large circle about A_1. The distribution of poles in the band of width $2\alpha R$ containing A_2 and A_3 is assumed to be uniform along the band, but the distribution across the band must be found. Divide the large circle about A_1 into rings of width $Rd\beta$. Consider one of radius $R\beta$. It contains approximately $2\pi R\beta Rd\beta k$ poles, the error being $\frac{1}{2}$ of one percent for α equal to 10°. Twice this number of associated poles are spread uniformly along 180° arcs that fall within a band of width $2\beta R$ (see Fig. 1). These arcs produce a uniform distribution of the poles. Considering every value of β, it may be seen that the distribution across the band, to a good approximation, is linear, varying from zero for β equal to α to a maximum at the center of the band. It may be represented by $a(\alpha-\beta)$ and must satisfy the following equations, the integration being carried out over $\frac{1}{2}$ the band along a 90° arc:

$$\pi\alpha^2R^2k=2\int_0^{\alpha}a(\alpha-\beta)R\pi/2\ Rd\beta.$$

$R\pi/2\ 2\alpha R$ gives the approximate area wherein one-half the associated poles fall. Consequently, $a=2k$.

We must now obtain all the poles in the circle about A_2 in order to calculate $P(G)$. Consider an arc where the density is $2k(\alpha-\beta)$. To a good approximation the circle intercepts a length of arc equal to $2[(R\alpha)^2-(R\beta)^2]^{\frac{1}{2}}$. Therefore,

$$P(G)=\frac{2\int_0^{\alpha}2k(\alpha-\beta)2[(R\alpha)^2-(R\beta)^2]^{\frac{1}{2}}Rd\beta}{2\int_0^{\alpha}2k(\alpha-\beta)\pi/2\ R\ Rd\beta}$$

$$=2\alpha(1-4/3\pi)=1.15\alpha.$$

The remaining poles of set G near A_3 are distributed within an approximate square of dimensions $2\alpha R$ in a rather complicated way. Consequently, the integration required to obtain $P(H)$ would be difficult to carry out. However,

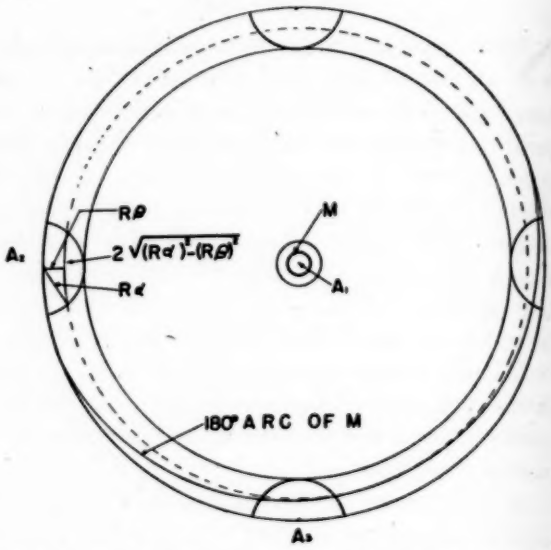

FIG. 1. Stereographic projection of the plot of set F together with diagrams used in calculating $P(G)$.

we can say definitely that

$$\frac{\pi\alpha^2R^2}{(2\alpha R)^2}<P(H)<1,$$

or $P(H)$ is between $\pi/4$ and 1. Substituting the values for $P(F)$, $P(G)$, and the limits of $P(H)$ in the equation for $P(E)$, we find that

$$1.35\alpha^3<P(E)<1.72\alpha^3.$$

Although this range is sufficiently accurate for most applications, a more definite value of $P(E)$ may be obtained as follows:

$$\text{write } P(E)=P(F)P(I).$$

The set I is to be obtained from F in such a way that all poles fall within the circles about

A_2 and A_3. Since a lengthy derivation is needed in this case, only the result will be given, namely:

$$P(I)=\frac{32}{\pi^2}(1-\sqrt{\tfrac{1}{2}})\alpha=0.95\alpha$$

(as before, approximations have been made for length of arc). Substituting the values for $P(F)$ and $P(I)$, we get:

$$P(E)=1.43\alpha^3. \qquad (1)$$

As an illustration, Eq. (1) gives $P(E)=1/1140$ for α equal to 5°. In this case, the maximum number of states (N) is 1140.

PROBABILITY OF OCCURRENCE OF AN ORIENTATION IN ONE OF SEVERAL GIVEN STATES

Case I: Two states, K and L, in which all orientations are uniquely classified. (We may have overlapping circles in this case, but one of the three circles for one state must not overlap a circle of the second state). The probability that an orientation will occur in state K by chance is:

$$P(K)=1.43\alpha^3.$$

The probability that it will occur in state L is also $1.43\alpha^3$. The probability that it will occur either in state K or L is:

$$P(K \text{ or } L)=P(K)+P(L)=2\times 1.43\alpha^3.$$

Case II: Two states, S and T, overlap so that a unique classification of all orientations is impossible and some have to be described as being in states S and T. In this case, we should think of each as a state only when it is alone; together, we should think of them as being less than two states.

The probability that an orientation will occur either in state S or T by chance is less than the value obtained for Case I. We note that if S and T are allowed to merge, the probability drops from $2\times 1.43\alpha^3$ toward $1.43\alpha^3$, corresponding to a transition from two states (Case I) toward one state. A correction owing to the amount of overlapping area might be made for the circle having the smallest amount of overlap, but the non-uniform distribution of poles within a circle for a state would make this correction difficult to determine. Nevertheless, there would be a number f such that $P(S \text{ or } T)=f1.43\alpha^3$, and $1<f<2$.

Case III: When data for standard twin groups through three generations are plotted on a stereographic net,[3] it is found that no overlapping occurs among the 52 orientations until α becomes equal to about 2½°. The probability that a new crystal will occur in one of the 52 states, when all N states are equally likely, is:

$$P=52/N=52P(E)=52\times 1.43\left(\frac{2.5\pi}{180}\right)^3$$

$$=0.00619=1/161.$$

When 5° is taken for α, the number of states N is 1140, but the maximum number of states allowed on the basis of a mechanism that produces only three orders of twins is 52. Because of overlap, there is some degeneracy among the states and their effective number is reduced as pointed out in Case II. Consequently,

$$P<\frac{52}{1140} \quad \text{or} \quad \frac{1}{22}.$$

This result means that recrystallization without any mechanism or preferment should not result in a 5° type fit more than once in every 22 times if we restrict a fit to one of the 52 first-, second-, and third-order twins. A mechanism of recrystallization of this degree of complexity should not be hard to test experimentally.[4]

PROBABILITY OF OCCURRENCE OF ORIENTATIONS FOR VARIOUS TYPES OF GROUPINGS

Consider a group of random orientations existing in N states, $H_1, H_2, H_3 \cdots H_N$ ($N=1/P(E)$). Select n orientations ($n \leqslant N$) in an independent manner and suppose that these fall into l states in such a way that m_1 belong in one state, m_2 in another state, etc., up to m_k in a kth state ($m \geqslant 2$), and one each belongs in the remaining $l-k$ states. Suppose further that some of the k states contain an identical number of orientations. Designate the number of equal size groups by r_1, r_2, etc. We ask, "What is the probability of occur-

[3] Data on four generations of twins have been obtained at the Pittsfield Works' Laboratory and are now being prepared for publication.

[4] In this connection, considerable data have already been accumulated at the Pittsfield Works' Laboratory of the General Electric Company.

rence of such a group?" The group in this case may be represented as follows:

$$H_{x(1)}^{m(1)}H_{x(2)}^{m(2)}H_{x(3)}^{m(3)}\cdots H_{x(k)}^{m(k)}H_{x(k+1)}^{1}H_{x(k+2)}^{1}\cdots H_{x(l)}^{1},$$

where the superscript* gives the number of orientations in a particular state. We do not, however, specify the states in advance. As an example, consider the case of throwing three dice to obtain a pair. In one throw we use three orientations (n) out of six (N), but to get a pair one orientation must occur twice. Therefore, l equals 2, k equals 1, and m_1 equals 2. Figure 2 gives an illustration of a more general orientation group

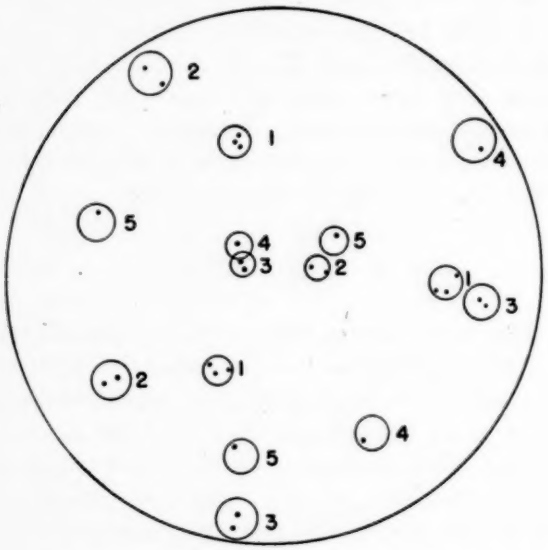

FIG. 2. Stereographic diagram showing states 1, 2, 3, 4, and 5 containing 3, 2, 2, 1, and 1 orientations, respectively. α is 5°. $N \cong 1000$.

existing in five states with n, l, k; m_1, m_2, m_3; r_1 and r_2 equal to 9, 5, 3; 3, 2, 2; 2 and 0, respectively.

To obtain the probability of occurrence for the general case, consider first any group of k states of the following type:

$$H_{x(1)}^{m(1)}H_{x(2)}^{m(2)}H_{x(3)}^{m(3)}\cdots H_{x(k)}^{m(k)}.$$

The number of combinations of N states taken k at a time for a group where

$$m_1=m_2=m_3\cdots=m_k \quad \text{is} \quad \frac{N!}{(N-k)!k!}.$$

The number of combinations, however, is increased if all the m's are not the same. Suppose one of the m's occurs r_1 times, another r_2 times, etc. The number of unique arrangements[2b] of m_1, $m_2\cdots m_k$ is the factor needed. It is $k!/r_1!r_2!\cdots$. The number of combinations leading, therefore, to a

$$H_{x(1)}^{m(1)}H_{x(2)}^{m(2)}H_{x(3)}^{m(3)}\cdots H_{x(k)}^{m(k)}$$

type group is:

$$\frac{N!}{(N-k)!k!}\times\frac{k!}{r_1!r_2!\cdots} \quad \text{or} \quad \frac{N!}{(N-k)!r_1!r_2!\cdots}.$$

We may also write combinations for the $l-k$ states that are occupied by only one orientation each. Since k states of the N have been used in forming the group $m_1m_2\cdots m_k$, there are $N-k$ remaining to select from to get the $l-k$ singly-occupied states. The number of combinations is:

$$\frac{(N-k)!}{[N-k-(l-k)]!(l-k)!} \quad \text{or} \quad \frac{(N-k)!}{(N-l)!(l-k)!}.$$

Any one of the combinations containing n orientations could have been selected in a number of unique ways, namely,

$$\frac{n!}{m_1!m_2!\cdots m_k!}.$$

Therefore, the total number of ways to form a specified type of group is:

$$\frac{N!}{(N-k)!r_1!r_2!\cdots}\times\frac{(N-k)!}{(N-l)!(l-k)!}\times\frac{n!}{m_1!m_2!\cdots m_k!}.$$

Since the total number of ways to make n selections is N^n, the probability of getting the specified type of group is:

$$P^k_{m(1)m(2)\cdots m(k)} = \frac{N!n!}{m_1!m_2!\cdots m_k!(N-l)!(l-k)!N^nr_1!r_2!\cdots}. \quad (2a)$$

When $l=n$, this equation reduces to:

$$P^0=\frac{N!}{(N-n)!N^n},$$

an expression giving the probability that all n orientations occur in different states.

* In superscripts and subscripts m_1 is written as $m(1)$; m_2 as $m(2)$; m_3 as $m(3)$; etc.

TABLE I. Probability of occurrence of orientations for various types of grouping.

k	Type of group m_1	m_2	m_3	m_4	Probability from Eqs. (2a) and (2b)	Probability
0					$P^0=\dfrac{N!}{(N-n)!N^n}$	P^0
1	2				$P_2^1=\dfrac{P^0}{2!}\dfrac{(N-n)!n!}{(N-n+1)!(n-2)!}$	$P_2^1=\dfrac{P^0}{2}\dfrac{n(n-1)}{N-n+1}$
2	2	2			$P_{22}^2=\dfrac{P^0}{2!2!}\dfrac{(N-n)!n!}{(N-n+2)!(n-4)!2!}$	$P_{22}^2=\dfrac{P_2^1}{4}\dfrac{(n-2)(n-3)}{N-n+2}$
1	3				$P_3^1=\dfrac{P^0}{3!}\dfrac{(N-n)!n!}{(N-n+2)!(n-3)!}$	$P_3^1=\dfrac{P_2^1}{3}\dfrac{(n-2)}{N-n+2}$
3	2	2	2		$P_{222}^3=\dfrac{P^0}{2!2!2!}\dfrac{(N-n)!n!}{(N-n+3)!(n-6)!3!}$	$P_{222}^3=\dfrac{P_2^1}{8}\dfrac{(n-3)(n-4)(n-5)}{N-n+3}$
2	2	3			$P_{23}^2=\dfrac{P^0}{3!2!}\dfrac{(N-n)!n!}{(N-n+3)!(n-5)!}$	$P_{23}^2=\dfrac{P_3^1}{2}\dfrac{(n-3)(n-4)}{N-n+3}$
4	2	2	2	2	$P_{2222}^4=\dfrac{P^0}{(2!)^4}\dfrac{(N-n)!n!}{(N-n+4)!(n-8)!4!}$	$P_{2222}^4=\dfrac{P_{23}^2}{32}\dfrac{(n-5)(n-6)(n-7)}{N-n+4}$
1	4				$P_4^1=\dfrac{P^0}{4!}\dfrac{(N-n)!n!}{(N-n+3)!(n-4)!}$	$P_4^1=\dfrac{P_3^1}{4}\dfrac{(n-3)}{N-n+3}$
1	5				$P_5^1=\dfrac{P^0}{5!}\dfrac{(N-n)!n!}{(N-n+4)!(n-5)!}$	$P_5^1=\dfrac{P_4^1}{5}\dfrac{(n-4)}{N-n+4}$
At least two in one state					$1-P^0$	$1-P^0$

TABLE II. Values of $P^k_{m(1)m(2)\cdots}$ for $N=1000$.

k	Type of group m_1	m_2	m_3	m_4	$n=20$ $P^k_{m(1)m(2)\cdots}$	$n=20$ $\Sigma P^k_{m(1)m(2)\cdots}$	$n=10$ $P^k_{m(1)m(2)\cdots}$	$n=10$ $\Sigma P^k_{m(1)m(2)\cdots}$	$n=5$ $P^k_{m(1)m(2)\cdots}$	$n=5$ $\Sigma P^k_{m(1)m(2)\cdots}$
0					0.8259284	0.8259284	0.9558606	0.9558606	0.9900350	0.9900350
1	2				0.1599657	0.9858941	0.0434044	0.9992650	0.0099401	0.9999751
2	2	2			0.0124616	0.9983557	0.0006125	0.9998775	0.0000149	0.9999900
1	3				0.0009773	0.9993330	0.0001166	0.9999941	0.0000099	0.9999999
3	2	2	2		0.0005070	0.9998400	0.0000030	0.9999971	—	
2	2	3			0.0001352	0.9999752	0.0000024	0.9999995	9.9×10^{-9}	
4	2	2	2	2	0.0000117	0.9999869	4.5×10^{-9}		—	
1	4				0.0000042	0.9999911	0.0000003	0.9999998	5.0×10^{-9}	
1	5				1.4×10^{-6}		3.2×10^{-10}		9.9×10^{-13}	
At least two in one state					0.1740716		0.0441394		0.0099650	

If we substitute P^0 in Eq. (2a), we get:

$$P^k_{m(1)m(2)\cdots m(k)}=\frac{P^0(N-n)!n!}{m_1!m_2!\cdots m_k!(N-l)!(l-k)!r_1!r_2!\cdots}. \quad (2b)$$

Successive applications of Eq. (2b) give the probabilities listed in Table I.

If a small value of n is taken, it can be shown readily for any N whatsoever that the sum of the probabilities for all possible groups is one.

For instance,

$$P^0+P_2^1+P_{22}^2+P_3^1+P_{32}^2+P_4^1+P_5^1=1,$$

when n is 5 and $N \geqslant 5$.

The values of $N!$, etc., in the expression for P^0 may be obtained by using Stirling's formula. If n is small, however, it is easy to compute P^0 directly with the aid of either logarithms or an electrical computing machine.

Table II gives some numerical results on probabilities for one value of N and three values of n.

ACKNOWLEDGMENTS

The writer wishes to express his appreciation to Dr. David Harker and Mr. Ralph E. Wareham for their reading of this paper. He is also indebted to the Calculating Section of the Pittsfield Works for aid in the numerical calculations.

PHYSICAL REVIEW VOLUME 66, NUMBERS 7 AND 8 OCTOBER 1 AND 15, 1944

The Photoconductivity of Sodium Chloride in the Far Ultraviolet

J. N. FERGUSON, JR.*
Cornell University, Ithaca, New York
(Received August 7, 1944)

Measurements have been made on the wave-length dependence of the photoconductivity of an uncolored sodium chloride crystal in the far ultraviolet region of 900 to 1350 angstroms in order to test the theories of Peierls, Frenkel, and Mott regarding the existence of discrete energy levels, called exciton levels, lying just below the conduction bands. A region was found between 1300 and 1350 angstroms in which there was large absorption but very small if any photoconductivity or external photoelectric effect indicating the existence of exciton states. At wave-lengths shorter than 1300 angstroms both the photoconductivity and the photoelectric effect began to set in and these showed a doublet structure. The photoconductivity rises abruptly at the second absorption peak in the absorption spectrum and shows structure which locates energy levels above the ground state. The experiments of Podubnij on the iodides are a contradiction to this experiment inasmuch as the alkali halides should reasonably be expected to behave similarly.

INTRODUCTION

PEIERLS,[1] Frenkel,[2] and Mott[3] have introduced the concept of non-photoconductive absorption in crystalline insulators because of the possible existence of discrete energy levels lying just below the conduction bands. The withdrawal of an electron by light of quantum energy corresponding to the distance AB, Fig. 1, leaves a hole in the filled band below A. This electron, termed an "exciton" by Frenkel, is then thought to be in the electric field of its positive hole, giving rise to discrete energy levels. Absorption of light of quantum energy corresponding to the distance AE would be a photoconductive absorption. A test for the existence of these exciton states would be to compare the dependence of photoconductivity and absorption on wave-length for an alkali halide crystal. Experimental data on the photoconductivity of pure alkali halides are very meager. To the author's knowledge the only previous work is by Podubnij[4] in which potassium and sodium iodide were illuminated over a wave-length range of 2500 to 1900 angstroms. The data showed photoconductive currents which depend in the same way on wave-length as does the absorption, and Podubnij concluded that all the absorbed energy is used to produce photoconductivity. There is no evidence that any care was taken to separate the primary photoconductive currents from secondary currents which Gudden and Pohl[5] have shown must be done in order to obtain interpretable results.

---

* Now at Corning Glass Works, Corning, New York.
[1] R. Peierls, Ann. d. Physik **13**, 905 (1932).
[2] J. Frenkel, Physik. Zeits. Sowjetunion **9**, 158 (1936).
[3] N. F. Mott, Trans. Faraday Soc. **34**, 500 (1938).
[4] Podubnij, J. Exper. and Theo. Phys. **8**, 410 (1939).
[5] Hughes and DuBridge, *Photoelectric Phenomena* (McGraw-Hill Book Company, Inc., New York, 1931), p. 291.

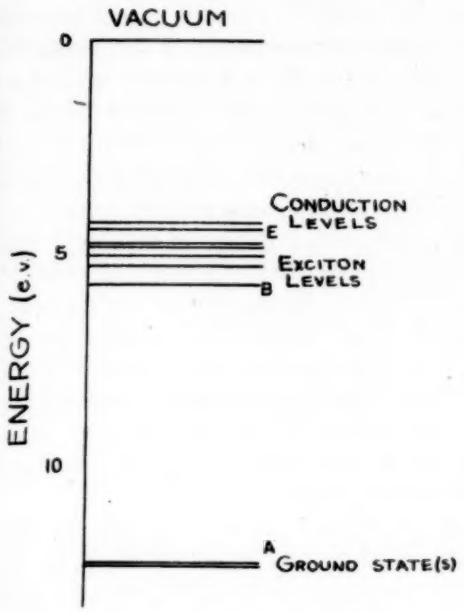

FIG. 1. Energy level scheme.

In view of the meager information about the photoconductivity of pure alkali halides, in addition to the desirability of obtaining evidence to confirm the existence of the exciton states introduced by Peierls, Frenkel, and Mott, it was considered worth while to attempt measurements of the photoconductivity of an easily produced and easily handled salt such as sodium chloride.

APPARATUS AND EXPERIMENTAL PROCEDURE

Illumination of the crystal was accomplished by mounting the crystal, as shown in Fig. 2, in a vacuum ultraviolet normal incidence spectrograph. The grating could be rotated to give various wave-lengths on the crystal. The source of illumination was a capillary air discharge tube. An electrically controlled shutter was placed at F between the exit slit and the grating so that the crystal could be illuminated for short fixed periods of time.

In order to measure relative intensities of the spectral lines used, a windowless silver cathode photoelectric cell was placed at E on the side of the crystal holder away from the grating. The crystal could be moved into or out of the beam from outside the vacuum.

The electric circuit for measuring the quantity of charge released by light flashes and for measurements of currents in the photoelectric cell was the Barth type electrometer tube circuit whose sensitivity was 310,000 millimeters per volt drop across the input resistor.

Thin single crystals of sodium chloride were produced by cleaving artificial crystals of high purity.[6] In order to decrease the possibility of surface conduction being measured as volume conductivity, the electrodes were behind the crystal and the crystal was mounted between them and the exit slit of the spectrograph. The technique developed by Gudden and Pohl[7] for elimination of secondary effects of polarization and decrease of resistance owing to release of electrons in the crystal was followed. This consisted

[6] Harshaw Chemical Company, Cleveland, Ohio.
[7] Hughes and DuBridge, reference 5, p. 289.

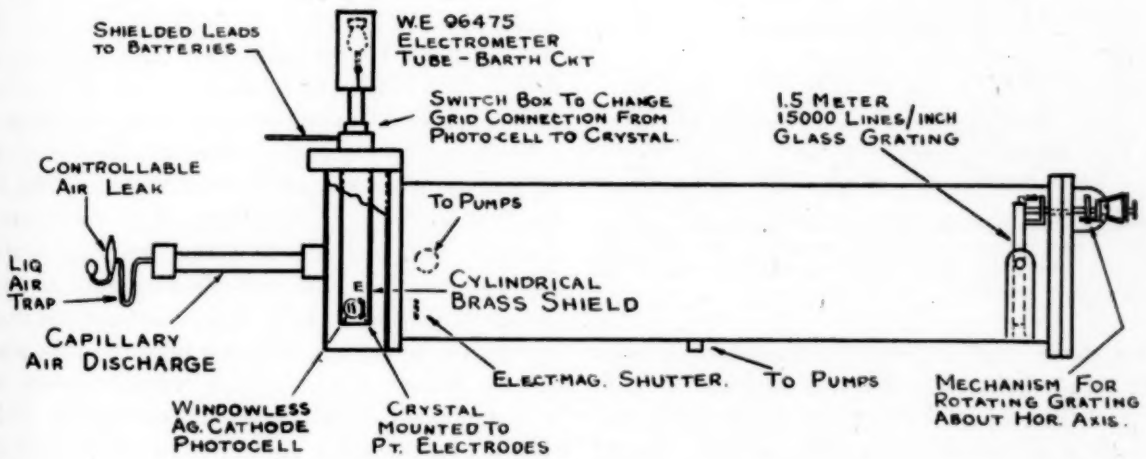
SHIELDED LEADS TO BATTERIES
W.E. 96475 ELECTROMETER TUBE - BARTH CKT
SWITCH BOX TO CHANGE GRID CONNECTION FROM PHOTO-CELL TO CRYSTAL
1.5 METER 15000 LINES/INCH GLASS GRATING
CONTROLLABLE AIR LEAK
TO PUMPS
LIQ AIR TRAP
E
CYLINDRICAL BRASS SHIELD
CAPILLARY AIR DISCHARGE
ELECTMAG. SHUTTER.
TO PUMPS
MECHANISM FOR ROTATING GRATING ABOUT HOR. AXIS.
WINDOWLESS AG. CATHODE PHOTOCELL
CRYSTAL MOUNTED TO PT. ELECTRODES

FIG. 2. Schematic diagram of apparatus.

in measuring all crystal photo-effects ballistically with low light intensity flashes and with a fairly low applied electrical field intensity in the crystal. The procedure was to set a spectral line on the slits by observing the photo-cell currents with the crystal out of the beam. After the intensity was observed to be steady within ten or fifteen percent, and it was usually much steadier than this, the shutter was closed; the crystal slid into position; and the electrometer tube grid connected to the crystal. The shutter was then opened for a fixed period of about one second by a timing device. The ballistic reading thus obtained was found to be made up of both external photoelectric currents as well as photoconductivity currents. These were separated in the following manner: The reading just described was made with a retarding potential, small compared to the potential drop across the crystal, between the

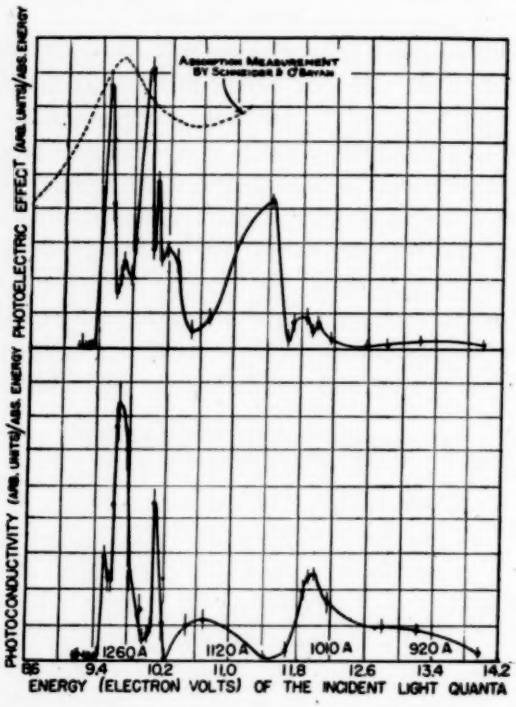

FIG. 3. Plot of photoconductivity current and photoelectric effect current in arbitrary units per unit absorbed energy as a function of the energy in electron volts of the incident light quanta.

brass shield, Fig. 2, and the crystal electrodes. The value of this potential was adjusted so as to be sure that all photoelectrons ejected from the crystal were collected by the positive crystal electrode. The reading i_a is the photoelectric effect of the crystal i_e, plus the photoconductivity i_c. To obtain measurement of the photoelectric effect alone, a small potential was applied across the crystal in a direction opposite to the large potential in the above measurement. This small potential was to insure that photoconductivity would not contribute appreciably to the current. A large accelerating potential was next applied, and a ballistic reading was made as before. This is the current i_e. Saturation of the current thus measured with accelerating voltage indicated that most of the photoelectrons were being measured. No measurable deflections were obtained at any wave-length if the crystal were removed from the crystal holder and normal operating conditions established.

Sources of error in this experiment are the following: (1) Unsteadiness of the light source. The light source gave two sources of unsteadiness: one owing to the nature of the method of excitation of the discharge; the other to non-uniform rate of leak of air into the tube. The discharge was remarkably steady, however, and the photoelectric cell showed that over a period of half an hour the average intensity taken over the period of the galvanometer system never varied by more than 25 percent and was usually 10 percent or less. (2) Non-monochromatism of the light flashed on the crystal. This was owing to three causes: first was the necessity of a broad slit width of approximately 0.5 mm in order to furnish enough light to get substantial galvanometer deflections. The second was the scattering of light into the slit. This was of course minimized in the customary ways by baffles and black walls. The third cause was the presence of overlapping orders of the spectrum. It is thought that the second-order intensity is small compared to the first-order intensity. Some evidence of this is that the photocell showed less than one-third as much current for the line 1085A in the second order as it did in the first order. (3) Unsteadiness of the electrometer tube. The proximity of a discharge tube to the electrometer tube brought about magnetic and electrical pick-up. These were so minimized that during the final runs reported here the tube was steady enough that readings could be made to plus or minus half-centimeter deflection. This was in general never greater than ten percent of the deflection.

RESULTS AND DISCUSSION

Overlapping data taken on separate days are plotted in Figs. 3 and 4. The photoconductivity current i_c was obtained from the data by the relation

$$i_a = i_e + i_c.$$

Assuming that the yield of the photoelectric cell, i.e., the number of electrons released per unit incident energy, is a slowly varying function over the spectral region studied, one can take the photoelectric cell reading i_p to be proportional to the intensity of the light and compare intensities at different wave-lengths. This assumption was made and the yields in arbitrary electrical units per unit absorbed light energy were calculated, the crystal photoelectric effect yield being i_e/i_p, and the photoconductivity yield being i_c/i_p. It is noticed in Figs. 3 and 4 that there is a region of large absorption in which there is no measurable photoconductivity or photoelectric effect (even though the photo-cell showed considerable light intensity in this region), but that these both set in sharply and almost at the same wave-length. These facts are in agreement with the exciton theory; the absorbed energy in the region in which we have neither photoconductivity nor external photoelectric effect is used in the formation of excitons. Curves for both the photoconductivity and photoelectric effect show doublet structure. Since the ground state of chlorine is a doublet of separation 0.11 electron volt, the first four peaks in the photoconductivity curve can be accounted for by assuming two narrow unfilled energy bands at 9.53 ± 0.02 and 10.06 ± 0.02 electron volts above the higher of the doublet ground states. The other peaks and region of continuum are explained by locating a band beginning at about 10.5 electron volts and another at about 11.6 electron volts above the higher of the ground state levels. Since the data were taken at larger energy intervals in the shorter wave-length regions, location of the latter two levels is approximate and there may be other levels as well. In comparing the two curves of photoelectric effect *versus* energy of the incident light quanta

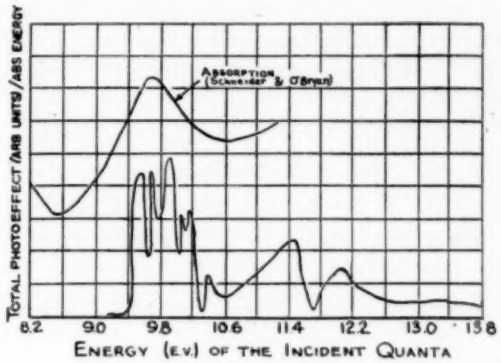

FIG. 4. The total photoelectric effect in arbitrary units per unit absorbed light energy as a function of the energy of the incident light quantum.

and photoconductivity *versus* the energy of the incident light quanta, it is noticed that peaks in the former curve occur where troughs of the latter curve occur, and vice versa. This would be expected if one realizes that any electron, given energy equal to or more than a certain minimum amount, can be ejected from the crystal as a photoelectron and is so ejected unless the value of the energy given is the correct amount to place the electron in a photoconductivity band. In that case the electron is a photoconductivity electron which moves through the crystal in the general direction of the field and the photoelectric current is small for the corresponding wave-length.

ACKNOWLEDGMENT

The author takes pleasure in expressing his appreciation for the advice and encouragement of Professor L. P. Smith during the course of this work.

Letters to the Editor

PROMPT publication of brief reports of important discoveries in physics may be secured by addressing them to this department. The closing date for this department is the third of the month. Because of the late closing date for the section no proof can be shown to authors. The Board of Editors does not hold itself responsible for the opinions expressed by the correspondents. Communications should not in general exceed 600 words in length.

New Researches in Gyromagnetism[1]

S. J. BARNETT

University of California, Los Angeles, California, and California Institute of Technology, Pasadena, California

September 20, 1944

AT the Strasbourg *Réunion sur le Magnétisme*,[2] where it devolved upon me to make the report on gyromagnetic phenomena, the President, Professor P. Weiss, said to me: "Your work has now become classical, but it must be extended to more substances and to the same substances in different states; and for this purpose it is necessary to increase the precision. And you are the man that has to do the work." This extensive program was, in fact, already underway, but it is urgent that many others should participate in it. It is difficult because the theoretically important differences between the gyromagnetic ratios are relatively small and the errors which have to be eliminated are seriously large. This investigation constitutes the initial part of the program, and consists of new and completely independent determinations, by means of the effect known as *magnetization by rotation*, or the *Barnett effect*, of the gyromagnetic ratios of a number of ferromagnetic substances. Although it is now certain that the work can be greatly improved after the end of the war, important confirmations of earlier work have already been achieved, and discrepancies have been reduced.

The general method of electromagnetic induction by which the effect involved was discovered in 1914 was used, but with numerous and important modifications. The sensitivity was greatly increased by using two galvanometers, the principal galvanometer acting as a fluxmeter and the other being actuated by it with the help of two photoelectric cells or, in about one-quarter of the work, a Moll relay. In most of the work, including all of that with the photoelectric cells, deflections were closely proportional to flux changes; and only this work is reported in the tables below. The magnetic rotors were identical with, or similar in construction to, those used by S. J. and L. J. H. Barnett in Washington, in magnetometer measurements on the same effect. The sources of error were reinvestigated and were largely eliminated, most of them by processes used in the earlier work. In all the new work the rotors were driven at exactly 60 r.p.sec. by small synchronous motors. In the greater part of the work the rotor axis was horizontal, and the rotor was driven from either end at will. It appears that only by driving from both ends can all of the errors be completely eliminated. In the latest work the axis was vertical, and the rotor was driven from above only. An important improvement consisted in making the sensitivity and gyromagnetic measurements simultaneously.

TABLE I. Gyromagnetic ratios (ρ) from new determinations ($e/m = 1.759$ e.m.u.).

Rotor	No. of sets	$\rho \times e/m$	Rotor	No. of sets	$\rho \times e/m$
Electrolytic iron III	15	1.025 ± 0.012	Cobalt I	16	1.072 ± 0.006
Electrolytic iron IV	12	1.031 ± 0.011			
Norway iron	4	1.028 ± 0.011	Hopkinson's iron-nickel alloy	19	1.019 ± 0.016
Mean for soft iron	31	1.028 ± 0.002	Permalloy	6	1.053 ± 0.006
Steel III	10	1.039 ± 0.008	Cobalt-iron	22	1.029 ± 0.009
			Cobalt-nickel	13	1.080 ± 0.020
Nickel I	17	1.054 ± 0.028	Heusler alloy	21	0.989 ± 0.006

TABLE II. Gyromagnetic ratios (ρ) of ferromagnetic substances. Values of $\rho e/m$ from the three extensive investigations in the author's laboratories ($e/m = 1.759$ e.m.u.).

Material investigated: Investigation	Effect	Soft iron	Steel	Nickel	Hiper-nik	Hopk. alloy	Perm-alloy	Cobalt	Cobalt-iron	Cobalt-nickel	Heusler alloy
I	Barnett	1.049	1.047	1.031	—	1.016	1.054	1.070	1.067	1.068	1.011
II	Einstein-de Haas	1.032	1.038	1.051	1.051	1.023	1.046	1.085	1.025	1.076	—
III	Barnett	1.028	1.039	1.054	—	1.019	1.053	1.072	1.029	1.080	0.989
Mean I and III	Barnett	1.038 ±0.010	1.043 ±0.004	1.042 ±0.010	—	1.018 ±0.002	1.054 ±0.000	1.071 ±0.001	1.048 ±0.019	1.074 ±0.006	1.000 ±0.011
Mean of II and mean I and III	Both	1.035 ±0.003	1.040 ±0.002	1.046 ±0.004	—	1.017 ±0.001	1.050 ±0.004	1.078 ±0.007	1.036 ±0.012	1.075 ±0.001	—

The means of the gyromagnetic ratios obtained as indicated above are given in Table I. A *set* consisted of 48 (in some cases 64) observations made in a symmetrical manner and on a strict time schedule.

In Table II these values are compared with those obtained in Washington from an investigation of the same effect but by the magnetometer method, and with those (II) obtained in Pasadena from an elaborate investigation of the converse effect (Einstein-de Haas effect). For the purposes of this table all values for soft iron are combined into one mean, likewise all values for steel, etc.

For I the authors claimed only a mean error for *all the rotors* not greater than 2 percent. It has now long been clear, and the present investigation makes it still clearer, that the mean error for the *individual rotors* in the Washington work is less than this. The values in II are considered by the author the most precise values of the gyromagnetic ratios hitherto obtained, those for Permalloy and soft iron being estimated to be correct to one-half percent or less. In view of the difficulties involved the three investigations agree remarkably well. The large value for soft iron in I is owing to the inclusion of an excessive result for a certain rotor, electrolytic iron II, and is not confirmed by II, as electrolytic iron IV was cut from the same rod, and electrolytic iron III was cut from similar material. The large value for cobalt-iron in I is not confirmed by either II or III.

For important papers on the theory, which, so far as it has yet been developed, is in agreement with such values as are given above. See Gorter and Kahn[3] and C. J. Gorter;[4] also reports by Gorter and Kronig and R. Forrer[5] and the discussions by Kramers and others.[6] See also R. Forrer.[7]

[1] Report of a paper presented by invitation of the American Physical Society at its 1942 Los Angeles meeting. See Proc. Am. Acad. Arts Sci. 75, No. 5 (August, 1944).
[2] The *Rapports* of this Congress were printed in France by the International Institute of Intellectual Cooperation and the French Central Service for Scientific Research, but their general distribution has hitherto been prevented by the war.
[3] Gorter and Kahn, Physica 7, 753 (1940).
[4] C. J. Gorter, Phys. Rev. 60, 836 (1941).
[5] Gorter and Kronig and R. Forrer, Strasbourg Réunion (1939); see also reference 2.
[6] Kramers *et al.*, see reference 2.
[7] R. Forrer, Comptes rendus 207, 1390 (1938).

A Strong Infra-Red Radiation from Molecular Nitrogen in the Night Sky

JOEL STEBBINS, A. E. WHITFORD, AND P. SWINGS
Carnegie Institution of Washington, Mount Wilson Observatory, Pasadena, California

September 20, 1944

SOME four years ago, in the course of photoelectric measures of the colors of stars and nebulae, Stebbins and Whitford found an intense infra-red radiation from the night sky which is scores of, if not a hundred, times as strong as the ordinary persistent auroral line at 5577A. That the new radiation must come from the atmosphere was shown by its variation with zenith distance, by its usual progressive decrease through the night, and by its variability from night to night and from season to season.

Renewed interest was awakened on July 20, 1944 when the radiation was doubly strong and was varying rapidly, as much as 10 or 15 percent in five minutes. No marked magnetic disturbance occurred on that night. It was then suggested by Swings that the radiation was undoubtedly molecular in origin and very likely could be traced to the (0,0) band of nitrogen at about 10,450A.

With this suggestion as a guide the wave-length was determined with the photo-cell and various thicknesses of a glass filter with a steep gradient in the transmission curve across 10,000A. The filters were readily calibrated in the first-order solar spectrum of the 75-foot grating of the 150-foot tower telescope on Mount Wilson.

From accordant measures of the sky on several nights the following values were obtained:

From 1 filter, 10,427A, weight 1;
From 2 filters, 10,452A, weight 2;
Mean 10,444A±10A.

The assigned probable error does not allow for other nearby lines or bands, but the main radiation is so strong compared with everything else in the whole range from 3500A to 12,000A that probably no other radiation is effective.

The only possible identifications that suggest themselves for this night sky emission are:

(a) the forbidden $^1D-^1P$ multiplet of *NI*, λ10,407.3 −λ10,397.8;
(b) the (0,0) band of the first positive group of N_2.

The [*NI*] lines do not provide a satisfactory identification. Their mean wave-length λ10,403 is definitely shorter than that of the observed infra-red radiation. Moreover, their total intensity should only be 49 times that of the $^4S-^2P$ line at λ3466.4 which is absent from or hardly visible in the night sky. Hence the [*NI*] lines could not account for the intense radiation observed.

If emitted in the upper atmosphere, the (0,0) band of N_2 should have its intensity maximum somewhere near λ10,450. Failure to observe other strong N_2 bands indicates in the night sky a mechanism enhancing the (0,0) band relative to the other vibrational transitions. It is suggested that during the night N_2 molecules are brought in the level $v'=0$ of the $B^3\pi_g$ state in three-body recombinations $N+N+N_2 \rightarrow N_2+N_2^{exc}$.

The nitrogen atoms would be formed during the day through photo-dissociation of N_2 by solar radiation. The suggested mechanism can be considered only for the value $D(N_2)=7.38$ v of the heat of dissociation of N_2 which agrees almost exactly with that of the $v'=0$ level of the $B^3\pi_g$ electronic state. In fact, the energy $D=7.38$ v is just a little larger than $B^3\pi$, $v'=0$, but smaller than $B^3\pi$, $v'=1$.

This mechanism, which is similar to that suggested by Chapman for the [*OI*] night sky emission, implies the presence of a fairly large number of nitrogen atoms in the upper atmosphere. Although the photo-dissociation of N_2 by solar radiation is usually considered as very weak compared to that of O_2 it is by no means negligible. This question has been considered recently by Ta-You Wu.[1] To

the arguments in favor of the presence of nitrogen atoms in the upper atmosphere advanced by Wu one should add that N_2 molecules can be predissociated by radiation of wave-length shorter than approximately λ1400, according to a private communication of Dr. Herzberg.

The mechanism suggested here depends on the value $D(N_2)=7.38$ v; it would not be effective if the heat of dissociation $D(N_2)=9.76$ v, recently suggested by A. G. Gaydon[3] were adopted.

Details will appear in a forthcoming paper in the *Astrophysical Journal*.

[1] Ta-You Wu, Phys. Rev. **66**, 65 (1944).
[2] A. G. Gaydon, Nature **153**, 407 (1944).

An Observation on the Ionosphere During the Solar Eclipse of July 20, 1944

C. K. JEN, K. T. CHOW, Y. T. LO, AND C. C. KUAN
Radio Research Institute, National Tsing Hua University, Kunming, China
September 20, 1944

AN observation of ionospheric changes was made here during the solar eclipse of July 20, in the midst of a series of routine observations already in progress for some time. The eclipse was annular along the belt in southern Asia, including such places as India, Burma, and Indo-China. As viewed in Kunming, it was only partial, with a maximum of 76 percent in diameter. The transmitter and the receiver were both placed in a suburb of Kunming, separated by a distance of about two kilometers. The transmitter is of a conventional build for pulse transmission, having a peak power of about 500 w and an adjustable frequency range of 2–5 mc. The width of the pulse is about 20μ and there are 50 such pulses per second. The receiver is designed to accommodate a wide-band reception, so as to preserve the sharpness of the received pulses. The signal and its reflections are observed on the screen of an oscilloscope (with adjustable linear sweep and extra time calibrating arrangements). For the week preceding the eclipse, the transmitter was continuously operating on the frequency of 3 mc, and for the sake of continuity, it was decided to let it stay put on the eclipse day and a month thereafter. We have noted that for this frequency only the reflections from the *E* layer have been most consistently observed; the reflections from the *F* layer only begin to occur more prominently during late afternoons. The virtual height of the *E* layer has always remained sensibly constant around the value of 120 km, as much on the eclipse day as any other day. Since this is only a confirmation of the results obtained by many other experimenters, we do not think it worth while to show our results. In Fig. 1, there is shown the record of the amplitude variation of the *E* layer reflection on the eclipse day, as compared with that of the average of seven typical records taken during the days in the neighborhood of the eclipse.

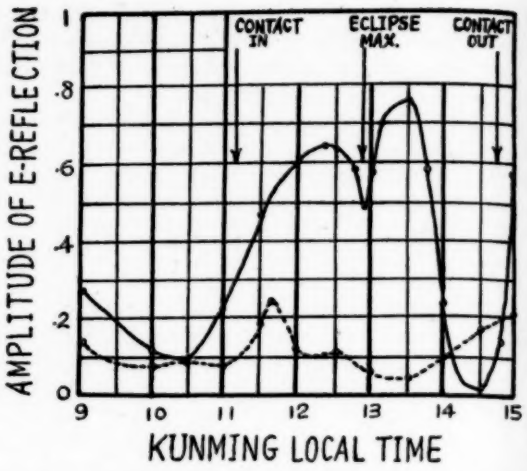

FIG. 1. Solid curve for amplitudes of *E* reflection on the eclipse day (July 20); dotted curve for amplitudes averaged from seven control days. Amplitudes of reflections are expressed as fractions relative to the received amplitudes of the ground wave.

The amplitude of each reflection represents the height to which the reflected pulse comes up most frequently (the echo wave has a tendency to fluctuate up and down) and is oftentimes smaller than a certain maximum peak that occurs less frequently. It is seen that on the average the amplitude of *E* reflection generally has a small peak just before noon but soon dies down thereafter and only begins to rise as the afternoon progresses. The record for the eclipse day is considered as remarkable in that it is very much different from the usual pattern and its rise and fall are fairly coincident with the beginning and end of eclipse. The general elevation of the reflected amplitude is most extraordinary. There is, further, a dip of amplitude, strangely coincident with the maximum eclipse. There was absolutely no wishful juggling of the result, because we had no idea which way to expect. These results evidently suggest a large decrease of absorption with respect to the *E* reflection during the solar eclipse. This diminishing of absorption may be caused by the changes occurring in the *E* layer itself as the result of decreased ionization. Or it may well be caused by a lower absorption layer that may control the amplitude variation. One may suppose that the rapid de-ionization and recombination processes that may take place in this hypothetical layer may be responsible for the decreased absorption. The constancy of the virtual height of the *E* layer and the extraordinary increase of the reflection amplitude seem to lend help to this hypothesis. But the mysterious middle dip is surely not readily explainable on this exceedingly crude model and there must be many other difficulties in explaining various other phenomena. The present experimental evidence is totally inadequate to decide between the two possibilities, but it is not impossible that the issue could be settled by studying more of the amplitude changes in the future.